

STIC-Biotech/ChemLib

80934

From: Collins, Cynthia
Sent: Friday, November 22, 2002 2:23 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request SN 09/868300

Please search, both prior art and interference, for SN 09/868300:

- 1) SEQ ID NO:7
- 1) SEQ ID NO:8

Thank You,

Cynthia Collins
Art Unit 1638
CM1, 9A12 (office) or 9E12 (mailbox)
(703) 605-1210

CRFE

12/17/98

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 12/4/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

Point of Contact
P. Sheppard

Telephone number: (703) 308-4499

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 19:10:05 ; Search time 2417 Seconds

(without alignments)
4375.529 Million cell updates/sec

Title: US-09-868-300-7

Perfect score: 653
Sequence: 1 gaatcgacacgagctccctt.....caaagatgagaccagcttg 653

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estbda:*
2: em_esthum:*
3: em_estlm:*
4: em_estlm:*
5: em_estlm:*
6: em_estlm:*
7: em_estlm:*
8: em_estlm:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlm:*
16: em_estlm:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hiv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501.4	76.8	529	10	AM004542 701932172
2	345.6	52.9	536	9	AI994093 701498662
3	345.2	52.9	765	13	BI204061 EST522101
4	335.6	51.4	712	12	BG596323 EST495001
5	326.6	50.0	823	10	BE05140 GA_F0003
6	322.2	49.3	745	10	AM218763 EST301243

7	320.8	49.1	557	9	AU236886	AU236886
8	314.2	48.1	566	14	BQ273330	BQ273330
9	311.6	47.7	441	17	BH634498	BH634498
10	311	47.6	452	14	R30079	R30079
11	276.2	42.3	1165	11	AY105313	AY105313
12	271.6	41.6	641	12	BG097619	BG097619
13	269.8	41.3	485	13	BW524990	BW524990
14	261.2	40.0	944	13	BI948508	BI948508
15	259	39.7	502	10	BE357810	BE357810
16	257	39.4	659	12	BG366429	BG366429
17	255.6	39.1	608	9	AL750043	AL750043
18	254.6	39.0	608	14	BQ415563	BQ415563
19	240.2	36.8	859	12	BF628880	BF628880
20	234.2	35.9	496	10	BE417248	BE417248
21	232.8	35.7	507	12	BE440848	BE440848
22	227.4	34.8	419	14	BQ273331	BQ273331
23	225.8	34.6	612	10	AM218764	AM218764
24	219	33.5	664	14	BQ803391	BQ803391
25	218	33.4	523	13	BM097653	BM097653
26	213.8	32.7	540	9	AJ470478	AJ470478
27	210.8	32.3	477	14	BQ753916	BQ753916
28	202.4	31.0	668	13	BI959690	BI959690
29	197	30.2	631	13	BI168895	BI168895
30	196.4	30.1	678	17	BH684219	BH684219
31	193.4	29.6	614	10	AV524167	AV524167
32	192.6	29.5	675	13	BI182893	BI182893
33	189.6	29.0	451	9	AU227890	AU227890
34	183.4	28.1	507	9	AI777460	AI777460
35	183	28.0	671	13	BI959716	BI959716
36	178.6	27.4	628	13	BM269335	BM269335
37	176.4	27.0	308	12	BE804809	BE804809
38	172.6	26.4	701	12	BG043358	BG043358
39	172	26.3	564	13	BI160655	BI160655
40	167.8	25.7	590	13	BM099343	BM099343
41	167.6	25.7	600	13	BM341015	BM341015
42	167	25.6	704	14	BQ624976	BQ624976
43	166.4	25.5	501	10	AM221667	AM221667
44	161.6	24.7	425	10	AV556381	AV556381
45	158.6	24.3	581	13	BI950438	BI950438

ALIGNMENTS

RESULT 1
AM004542
LOCUS 701932172 A. thaliana, mixed source Arabidopsis thaliana cDNA clone
DEFINITION 701932172. mRNA sequence.
ACCESSION AM004542
VERSION AM004542.1 GI:5851571
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 529)
AUTHORS Chen, J., Montyana, M., Chan, E., Mooney, M., Caroon, B., Gilliland, D., Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzozka, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argente, C., Shah, S., Noblitz, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression Microarray
Unpublished (1999)
CONTACT: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

FEATURES
source
Location/Qualifiers
1. .529
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="701932172"
/clone_lib="A. thaliana, mixed source"
/note="This sequence was obtained from a clone generated with a PCR product of the target gene."
BASE COUNT 136 a 114 c 118 g 160 t 1 others
ORIGIN

Query Match 76.8%; Score 501.4; DB 10; Length 529;
Best Local Similarity 98.9%; Pred. No. 1.5e-146;
Matches 526; Conservative 0; Mismatches 2; Indels 4; Gaps 2;

QY 101 CAAGTGAAGTGGCTCAATTAGTTCAGTAATCTGCGATTTCTTCAGATGGG 160
DB 1 CAAGTGAAGTGGCTCAATTAGTTCAGTAATCTGCGATTTCTTCAGATGGG 60
QY 161 CTGAAGTCAAACCGCTGCTCAAAATCATCTCATTTCTTGATTAACCTCTCTAC 220
DB 61 CTGAAGTCAAACCGCTGCTCAAAATCATCTCATTTCTTGATTAACCTCTCTAC 120
QY 221 CTCAGAAATGACTCTCTATACAGATGAAGATGCTGACCTGGAACAAGATTGGTG 280
DB 121 CTCAGAAATGACTCTCTATACAGATGAAGATGCTGACCTGGAACAAGATTGGTG 180
QY 281 ATCCGCTCTTATATATGAGCTCAGACGCTGGGCTGATTTATGATCATTTGCTCTTGT 340
DB 181 ATCCGCTCTTATATGAGCTCAGACGCTGGGCTGATTTATGATCATTTGCTCTTGT 240
QY 341 CTGCTAACACATTAGCCAAAGATTGGTGGTATGATTAATCTATTGACATGTATAG 400
DB 241 CTGCTAACACATTAGCCAAAGATTGGTGGTATGATTAATCTATTGACATGTATAG 300
QY 401 TAAGACATGGATTAACCAACCGTTGTTGTCACCGGCGATGAACACTTGTATGT 460
DB 301 TAAGACATGGATTAACCAACCGTTGTTGTCACCGGCGATGAACACTTGTATGT 360
QY 461 GGAACAAATCCCTTCAAGAAAGGACCTGCTGCTATGATTAACCTTGAATCAACCTTAA 520
DB 361 GGAACAAATCCCTTCAAGAAAGGACCTGCTGCTATGATTAACCTTGAATCAACCTTAA 417
QY 521 TTCTCCCATCAAGAAAGAACTGGCTGTGAGACTACGCTATGAGCGCAAT-GGCTGAG 579
DB 418 TTCTCCCATCAAGAAAGAACTGGCTGTGAGACTACGCTATGAGCGCAATGGCTGAG 477
QY 580 CCTTCTGATTTATTCACCTGTAGACTGTTCTGGGAGTCAACAGCTCGTA 631
DB 478 CCTTCTGATTTATTCACCTGTAGACTGTTCTGGGAGTCAACAGCTCGTA 529

RESULT 2
A1994093 536 bp mRNA linear EST 08-SEP-1999
LOCUS 701498662 A. thaliana, Ohio State clone set Arabidopsis thaliana
DEFINITION cDNA clone 701498662, mRNA sequence.
ACCESSION A1994093
VERSION A1994093.1 GI:5840998
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 536)
REFERENCE
AUTHORS Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guebler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argente, C., Shah, S., Nobrigha, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.
TITLE Arabidopsis thaliana Gene Expression Microarray

JOURNAL
COMMENT Unpublished (1999)
Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

FEATURES
source
Location/Qualifiers
1. .536
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="701498662"
/clone_lib="A. thaliana, Ohio State clone set"
/note="cDNA library was made from selected clones from the Arabidopsis thaliana Ohio State clone set."
BASE COUNT 137 a 118 c 109 g 172 t
ORIGIN

Query Match 52.9%; Score 345.6; DB 9; Length 536;
Best Local Similarity 84.9%; Pred. No. 1.9e-97;
Matches 387; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 197 ATTGCTGATTAACCTCTCTACCTCAGATGTGACTCTATACAGATGAAGATGAAT 256
DB 3 ATTTCCTCATTAACCTCTCTCCACAAAGATGACTCTGATATGATGAAGATGAAT 62
QY 257 GGTCTGTGGAACAAGATGGTATCCGCTTCTCATATGAGCTCAGACGCTGGGCTG 316
DB 63 GGTCTGTGGAACAAGATGGTATCCGCTTCTCATATGAGCTCAGACGCTGGGCTG 122
QY 317 ATGTATGATCATTTGCTCTTGTGCTATACACATTAGCCAAAGATTGGTGGTAT 376
DB 123 ATGTATGATCATTTGCTCTTGTGCTATACACATTAGCCAAAGATTGGTGGTAT 182
QY 377 GGTATATCTATTTGATCATGTATAGAGCATGGGATTAACCAACCGTTGTTGTTG 436
DB 183 GGTATATCTTCTGATCTTCAATTAACGAGCTTGGAGCATTAACCAACCGTTGTTG 242
QY 437 CACCGGCGAAGAACTTTGATGTGGAACAATCCCTTCAAGAAAGGACCTGTTTGC 496
DB 243 CTCACACTATGATTAATTTGATGTGGAACAATCCCTTCAAGAAAGGACCTGTTTGC 302
QY 497 TTGATGAACCTTGAATCAACCTTAATTCCTCCATCAAGAAAGAACTGGCTGTGAGACT 556
DB 303 TTGATGAACCTTGAATCAACCTTAATTCCTCCATCAAGAAAGAACTGGCTGTGAGACT 362
QY 557 ACGGTAATGGCGCAATGGCTGAGCTTCTGATTTATTCACCTGTTAAGCTGTTGGG 616
DB 363 ACGGTAATGGAGCTATGGCTGAGCCCTCTTATCTATTCACCTGTCAGACTCTTCTGGG 422
QY 617 AGTCACAGCTGTTAACAAGAGATGAACCAAGTT 652
DB 423 AGTCACAGCTGTTAACAAGAGATGAACCAAGTT 458

RESULT 3
B1204061 765 bp mRNA linear EST 11-JUL-2001
LOCUS B1204061
DEFINITION EST522101 cTOS Lycopersicon esculentum cDNA clone cTOS3N14 5' end,
mRNA sequence.
ACCESSION B1204061
VERSION B1204061.1 GI:14681785
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 765)
REFERENCE
AUTHORS van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R., Ronning, C. and Tanksley, S.

[illegible][illegible]

QY	485	ACCTTGTCTGCTTGATGAACTTGGAAATCCACCCTAATTCCTCCCATCAAGAAGAAACGG	544
Db	573	ACCTTATGTATGATGATGAGCTTGGAAATCTCTCTCATACCAACAGTTTCTTAAAGACTAG	632
QY	545	CCTGTGAGACTACGAGTATATGGCGCAATGGCGCTTCTCTCATATTTATTCACCTGTTA	604
Db	633	CTTGTGAGAAATATGGAACAGCGCGGTATGGCTGAAACCTTCTCATCTACTCAACTGTAA	692
QY	605	GACTGTCTGGGAGTAC 622	
Db	693	GACTCTCTATGAGTACAC 710	
RESULT 5			
BE055140			
DEFINITION	BE055140	823 bp	mRNA linear EST 07-MAR-2001
LOCUS	GA__Ea0033A06f	Gossypium arboreum 7-10 dpa fiber library	Gossypium
ACCESSION	BE055140		
VERSION	BE055140.2	GI:13246149	
KEYWORDS	EST.		
SOURCE	Gossypium arboreum.		
ORGANISM	Gossypium arboreum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Wing, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry		
TITLE	, D., Wood, T.C., Leslie, A. and Wilkins, T.A.		
JOURNAL	An integrated analysis of the genetics, development, and evolution		
COMMENT	of the cotton fiber unpublished (2000)		
	On Jun 8, 2000 this sequence version replaced gi:8382197.		
	Contact: wing RA		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Tel: 864 656 7288		
	Fax: 864 656 4293		
	Email: rwing@clemson.edu		
FEATURES	Seq primer: TAAATACGACTACTATAGG		
source	High quality sequence stop: 637.		
	Location/Qualifiers		
	1..823		
	/organism="Gossypium arboreum"		
	/strain="AKA"		
	/cultivar="8400"		
	/db_xref="taxon:29729"		
	/clone="GA__Ea0033A06f"		
	/clone_lib="Gossypium arboreum 7-10 dpa fiber library"		
	/tissue_type="Fibers isolated from bolls harvested 7-10		
	dpa"		
	/lab_host="E. coli"		
	/note="vector: pBR-CMW; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	225 a	166 c	191 g
ORIGIN			241 t
Query Match	50.0%;	Score 326.6;	DB 10; Length 823;
Best Local Similarity	73.5%;	Pred. No. 2.5e-91;	
Matches 430; Conservative	0;	Mismatches 154;	Indels 1; Gaps
QY	44	AGAAGATGATATGATGATGATACAGTACAGTACAAAGAAGCCCTGTAATCTTACTAGCTGCAA	103
Db	20	ATAGGAGATGCTTATAGGATCAATCTCAACCCAGAAACCCCGGTTTACTCGTGCCA	79
QY	104	GTCGAGATGCGCTTCAATTAAGTCAAGTCAATCTGCGCATGTTTCTCAGATGGCGTG	163
Db	80	GTCGAGTGTAGCTGCCATTAAGATTGGGGAATCTGCCATTCTTCTGATGGGCGAG	139
QY	164	AAGTCAAAAGCCGCTGCTTCAAAATCACTCTCAATTTGCTTGAATAAAGCTTCTCACTC	223
140	AAGTAAAGCAGTGTGCACGAAGGCTTCTTGGATTTCATTTGACATAGCATCACTTCTCTA	199	

Query	Match	49.3%	Score 322.2	DB 10	Length 745
QY	224	AGAAATGACACTCTCTATACATAGAAAGTATGATGGTGAACAAAGATTTGGTGATC	283		
Db	200	AGGATCTAAAGCTTTTACACTGATGTGGAGGAATGGTCTTAGTTGGCGAAATAGGTGACA	259		
QY	284	CCGTTCTTCATATCGACGCTCAGACGCTGGGCGTGATGTATGATCAATTCCTCTTGTCTG	343		
Db	260	GTGTCTCTTACATATGAGTGTGTGATGGGCGTGATATATATGATATTCGCCCAATGTGAC	319		
QY	344	CTTAAACATTTAGCCAGATTTGCTGGTGTTATGTGATTAATCTATTGCATGTATAGTAA	403		
Db	320	CAAAACACACTTGGCAGATTTGCTGGAGGATTAATGACAAATTTGTTAACTGTGTGCTGAC	379		
QY	404	GAGCATGGGATTAATGATGACCAACCGTTGTTGTTGGACCGCGGCGTAACACTTTGATGTGGA	463		
Db	380	GAGCATTTGGACATACAGCAAGCAATGTTTGTGGACCCAGCATATGACACTTTCAATGTGGA	439		
QY	464	ACAATTCCTTTACAGAACGGCACCCTTGTCTTGCTTGATGATGAACCTTGAATCACCCATAATTC	523		
Db	440	GCAACCCCTTTTCACAAAAAGCATCTCATGACAAATTTGATGAGTGTGTATTTCTTCATCC	499		
QY	524	CTTCCCATCAAGAAAGAACTGGCCTTGTGGAGACTACGGTAATGGCCCAATGGCTGACCTT	583		
Db	500	CCCCGTCTCCAA-AGACTAGCTGTGTGGGAGCTAATGGGAAACGCCCAATGGAGAAACCTT	558		
QY	584	CTCTGATTTATTCACACTGTATGACGCTTCTGGAGACTGCACACCTC	628		
Db	559	CTCTAATCCACTGACTGTAAATTAATTTCTTGAGTCAACGACTC	603		
RESULT 6					
LOCUS	AW218763				
DEFINITION	AW218763 tomato root during/after fruit set, Cornell University				
ACCESSION	AW218763				
VERSION	AW218763.1				
KEYWORDS	EST.				
ORGANISM	tomato.				
REFERENCE	Lycopersicon esculentum				
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;				
TITLE	Lycopersicon.				
JOURNAL	1 (bases 1 to 745)				
COMMENT	van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romling,C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D. Generation of ESTs from tomato root tissue Unpublished (1999)				
FEATURES	Contact: CUGI				
source	Clemson University Genomics Institute				
	Clemson University				
	100 Jordan Hall, Clemson, SC 29634, USA				
	Email: http://www.genome.clemson.edu/orders/index.html				
	5 prime sequence.				
	Location/Qualifiers				
	1..745				
	/organism="Lycopersicon esculentum"				
	/cultivar="TA496"				
	/db_xref="taxon:4081"				
	/clone="CLEX1E15"				
	/clone_lib="tomato root during/after fruit set, Cornell University"				
	/tissue_type="root"				
	/dev_stage="plants during and after fruit-set"				
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Tanksley; tissue supplied by Dave Garvin (USDA-ARS, Ithaca, NY 14850)."				
	(USDA-ARS, Ithaca, NY 14850).				
	155 c 170 g 224 t				

[illegible]

and XhoI was ligated to modified Lambda Φ C101 vector (Carinciuc et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/Plant/index_e.html) for further details.

```

/db_xref="taxon:3702"
/clone="RAFL15-30-H14"
/clone_lib="RAFL15"
/tissue_type="mixture of silique and flower"
/lab_host="DH10B"
/notes="Site_1: BamHI; site_2: SalI"
145 a 121 c 127 g 161 t 3 others
BASE COUNT
ORIGIN

```

Query Match	49.1%	Score 3207			
Best Local Similarity	82.6%	Pred. No. 1.3e-89;			
Matches	389;	Mismatches	80;	Indels	2; Gaps
Conservative					

[illegible]

113 TGCGTTCATTAGTTCAGTAATCTCTGCCAATGTTTCTCAAAATGGCAGAG 20
 148 TCGCTGCTATCAAAATCGGCAATCTGTGCCATTGCTTCACCGAATGGCGAGAGTCAGAG 20

173 CCGTCGCTCAAAATCATCTCTCAATTCTGTGATAAACCTTCTACCTCAGAAATGTA 26
 QY ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
 208 CCGTCGTACGAATCATCTCTACATTTCCTGATAAATCTCTCTCCACAAGAAGTGA

[illegible][illegible][illegible]

D_b 388 TAGGCACAGATTGCTGTGGCGCTTGGATAAATCCTCAATCCAAATCTT
4
O_y 413 ATTATAGCAAACCGTTGTTTGTCACCGCGCATGAACAC-TTGTATGTGAACAATCTT
5

Db 448 ACTATACCAACCACCTGTTGTGGTCCAGCATTGAATACTTTTGATGCGGAAACAATTC

Ov 472 TTCACAGAACGCACCTTGCTGCTGCTGATGAACCTTGSATCACCCCTAAT 522

Db 508 TTTCACTGAA-GGNATNTTTCTGCTGATGAAGCTGGGATCACCATTAT 55/

[illegible]

DEFINITION	Sao23106.0.11 8" 5' similar to TR:Q9SWE5 Q9SWE5 HAUCH
ACCESSION	P0273330

VERSION	BQ273330.1	GI:20456400
KEYWORDS	EST.	
SOURCE	soybean.	
ORGANISM	glycine max	
		starch:beta: Amylase
		Embryophyta; Tracheophyta

Ononitaceae; Ranunculaceae; Papilionaceae; Fabaceae; Rosidae; eurosids I; Fabales; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Eukaryota; Viridiplantae; Streptophyta; Charophyta; Embryophyta.

REFERENCE
AUTHORS

Glycine: 1 (bases 1 to 566)
Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Corvill, V.,
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Stentoe, M., Theising, B., Allen, M.,

Wyllie, T., Underwood, N., Deeprose, J.

Db	291	TATCTTACTAGCTGCAAGTGAAGTGGGCTTCATTTAAGTTCACTAATCTCTGCAATTG	232
QY	147	TTTCTCAGAAATGGGCTGAAGTCAAAAGCCGTCGCTTCAAAATCATCTCTCATTTTCGTTGA	206
Db	231	TTTCTCAGAAATGGGCTGAAGTCAAAAGCCGTCGCTTCAAAATCATCTCTCATTTTCGTTGA	172
QY	207	TAAACCTTCTTACTCGAAGATGTAGCTCTATTCACAGTGAAGTGAAGTGAAGTGTAGCTG	266
Db	171	TAAACCTTCTTACTCGAAGATGTAGCTCTATTCACAGTGAAGTGAAGTGAAGTGTAGCTG	112
QY	267	GAACAGATGGTGTATCCGCTTCATATTCAGAGCTCAGAGCTGGGCTGATGTTATGAT	326
Db	111	GAACAGATGGTGTATCCGCTTCATATTCAGAGCTCAGAGCTGGGCTGATGTTATGAT	52
QY	327	CATTGCTCTCTTGTCTGCTAAACACATTAGCCAAAG	360
Db	51	CATGGCTGATGCTGCTAGCGCATATGACCAAG	18
RESULT	10		
R30079			
LOCUS		452 bp	mRNA
DEFINITION		12684 lambda-PR2 Arabidopsis thaliana cDNA clone 164P177, mRNA	
ACCESSION		R30079	
VERSION		R30079.1	GI:936772
KEYWORDS		EST.	
SOURCE		thale cress.	
ORGANISM		Arabidopsis thaliana	
REFERENCE		Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta; Spermatophyta; Magnoliophyta: eudicotyledons; core eudicots; Rosidae; eustoids II; Brassicales; Brassicaceae; Arabidopsis.	
AUTHORS		Newman,T., deBurlin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Rezel,E., and Somerville,C.	
TITLE		Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones	
JOURNAL		Plant Physiol. 106, 1241-1255 (1994)	
MEDLINE		95148729	
COMMENT		On Apr 14, 1993 this sequence version replaced gi:785394.	
FEATURES		Source	
source		1. 452	
		/organism="Arabidopsis thaliana"	
		/strain="var Columbia"	
		/db_xref="taxon:3702"	
		/clone="164P177"	
		/clone="164P177"	
		/note="Vector: lambda zip-lox; Site_1: Sal; Site_2: Not; lambda PR2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRU's lambda zip-lox. The cDNA inserts were directionally cloned with Sal not arms using 0.11ug of primed cDNA."	
BASE COUNT		100 a 100 c 99 g 135 t 18 others	
ORIGIN			
Query Match		47.6%	Score 311; DB 14; Length 452;
Best Local Similarity		82.2%	Prod. No. 1.3e-86;
Matches 347;		Conservative 0;	Mismatches 75; Indels 0; Gaps 0;

QY	197	ATTTCGTTGATTAACCTTCCTACCTACGAATGACCTCTATATACATATGAATGAT	256
Db	3	ATTTCCTGATTAACCTCTCTCTCCACAAGANGTACTCTTATATCTGATGAAGATGAAT	62
QY	257	GCTTAGCTGGAACAAGATTGGTGAATCCGTTTCTTCATATGAGCTCAGACGTGGGCTG	316
Db	63	GGNCTAGCTGGAAACAAGATCGGTGATCTTCCTTTCATCATGAGCTTGAAGCTTGGGCTG	122
QY	317	ATGTTATGATCATCTTCCTTTGCTGCTTAACCATTTAGCCAAAGTTGCTGGGTAT	376
Db	123	ATGTTTATGATCATCTTCCTTTGCTGCTTAACCATCTTGAAGAAATTCGCTGGGCTTT	182
QY	377	GTGATTAATCTATGATCATGTTATGATTAAGATGGATTTATAGCAAAACGTTGTTGTTG	436
Db	183	GTNATTAATCTATGATCATGCTTCATATACGACCTTGGGACTATATACCAACCACTGTTTGCG	242
QY	437	CACCGCGCATGAACACTTTTGATGTTGGAACAATCCTTTTCACAGAAGCGACCTTGCTTTCG	496
Db	243	CTCCAGCTATGATTAATCTTTCATATGAGAACATCCTTTTCACATGAAGGATCTTTNGTCTC	302
QY	497	TTGATGAATCTTGAATCAACCTTAATTCCTCCCATCAACAAGAACTGGCTGTGAGACT	556
Db	303	TTGATGAATCTGGGGATCAACATTAATTCCTCATCAAGAAAGNGCTTCCTGTGGGGCT	362
QY	557	ACGGTAATGGCGCAATGGCTGAGCCTTCCTGATTTATTCACATGTTAGACTGTTCTGGG	616
Db	363	ACGGTAATGGAGCTATGGCTGAGCCCTCTCTTATCTATTCACATGTCAGACTTTTNTGGG	422
QY	617	AG 618	
Db	423	NG 424	
RESULT 11			
LOCUS	AY105313	1165 bp	mRNA linear HTC 25-MAY-2002
DEFINITION	Zea mays PC0119881 mRNA sequence.		
ACCESSION	AY105313		
VERSION	AY105313.1 GI:21208391		
KEYWORDS	HTC.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.		
AUTHORS	1 (bases 1 to 1165) Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.		
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes		
JOURNAL	Unpublished (2002)		
REFERENCE	2 (bases 1 to 1165)		
AUTHORS	Coe,E.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA		
FEATURES	Location/Qualifiers		
source	1..1165		
	/organism="Zea mays"		
	/db_xref="MaizEDB:636974"		
	/db_xref="taxon:4577"		
	/clone="PC0119881"		
	/clone_lib="Maize Mapping Project/Dupont Consensus Library"		
	/note="this sequence is part of a project of EST assemblies resulting from the application of public access to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"		
BASE COUNT	291 a 275 c 307 g 292 t		
ORIGIN			

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 19:19:30 ; Search time 98 Seconds
(without alignments)
2566.126 Million cell updates/sec

Title: US-09-868-300-7
Perfect score: 653
Sequence: 1 gaattcgacagagctcctt.....caaagatggaaccagctg 653

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 19255720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.6	20.2	383	10	US-09-878-574-5053 Sequence 5053, Ap
2	121.2	18.6	903	9	US-09-764-904-34 Sequence 34, Appl
3	121.2	18.6	903	10	US-09-764-860-293 Sequence 293, App
4	60.2	9.2	4503	10	US-09-070-927A-149 Sequence 149, App
5	57.6	8.8	281	10	US-09-878-574-6517 Sequence 6517, Ap
6	52	8.0	11842	9	US-09-764-904-126 Sequence 126, App
7	52	8.0	11842	10	US-09-764-860-1187 Sequence 1187, Ap
8	48.2	7.4	1218	10	US-09-974-300-984 Sequence 984, App
9	44	6.7	787	9	US-09-895-913A-65 Sequence 65, Appl
10	34.4	5.1	394	10	US-09-878-919-37 Sequence 37, Appl
11	33.2	5.1	387	10	US-09-878-574-1714 Sequence 1714, Ap
12	33.2	5.1	15567	12	US-10-047-676A-3 Sequence 3, Appl
13	33	5.1	1067	10	US-09-974-300-5426 Sequence 5426, Ap
14	33	5.1	4527	10	US-09-901-940-3 Sequence 3, Appl
15	32.4	4.9	398	10	US-09-983-965-5883 Sequence 5883, Ap
16	32.2	4.9	277	9	US-10-046-935-2165 Sequence 2165, Ap
17	32.2	4.9	692	10	US-09-764-869-1990 Sequence 1990, Ap
18	31.6	4.8	568	10	US-09-764-869-1991 Sequence 1991, Ap
19	31.6	4.8	578	10	US-09-764-869-384 Sequence 384, App

20	31.6	4.8	2000	9	US-09-938-842A-2966 Sequence 2966, Ap
21	31.2	4.8	1738	10	US-09-925-300-294 Sequence 294, App
22	30.8	4.7	574	10	US-09-864-761-12057 Sequence 12057, A
23	30.8	4.7	1188	10	US-09-867-550-2027 Sequence 2027, App
24	30.6	4.7	377	10	US-09-895-828-237 Sequence 237, App
25	30.6	4.7	2000	9	US-09-938-842A-3829 Sequence 3829, Ap
26	30.4	4.7	557	9	US-10-046-935-359 Sequence 359, App
27	30.4	4.7	5194	9	US-10-002-389-1 Sequence 1, Appl
28	30.2	4.6	374	10	US-09-878-574-3522 Sequence 3522, A
29	30.2	4.6	418	10	US-09-960-352-14362 Sequence 14362, A
30	30.2	4.6	1268	10	US-09-070-927A-557 Sequence 557, App
31	30.2	4.6	1272	10	US-09-808-483-7 Sequence 7, Appl
32	30.2	4.6	30013	10	US-09-764-877-3297 Sequence 3297, Ap
33	30	4.6	675	10	US-09-974-300-6755 Sequence 6755, Ap
34	30	4.6	32038	10	US-09-864-761-4628 Sequence 4628, App
35	29.8	4.6	396	10	US-09-864-761-21370 Sequence 21370, A
36	29.8	4.6	396	10	US-09-864-761-21370 Sequence 21370, A
37	29.8	4.6	475	10	US-09-867-701-2393 Sequence 2393, Ap
38	29.8	4.6	3293	8	US-08-910-386A-17 Sequence 17, Appl
39	29.8	4.6	6101	10	US-09-972-714-1 Sequence 1, Appl
40	29.6	4.5	325	9	US-09-920-455-6 Sequence 6, Appl
41	29.6	4.5	360	10	US-09-895-828-231 Sequence 231, App
42	29.4	4.5	346	10	US-09-070-927A-752 Sequence 752, App
43	29.4	4.5	427	9	US-09-920-455-153 Sequence 153, App
44	29.4	4.5	612	10	US-09-974-300-5525 Sequence 5525, Ap
45	29.4	4.5	1432	10	US-09-822-830A-394 Sequence 394, App

ALIGNMENTS

RESULT 1
US-09-878-574-5053
Sequence 5053, Application US/09878574
Patient No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(1540)B
CURRENT APPLICATION NUMBER: US/09/878, 574
PRIOR FILING DATE: 2001-12-21
PRIORITY APPLICATION NUMBER: 09/333, 535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 5053
LENGTH: 383
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-011-Q1-B1-D1
US-09-878-574-5053

Query Match 20.2%; Score 131.6; DB 10; Length 383;
Best Local Similarity 65.1%; Pred. No. 3.5e-33;
Matches 194; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY	40	GAGAGAAGATGATATGGAAGTGAAGTACAGCAAGCAAGCTGTACTTACTACT	99
DB	86	GAATGAGAAAACCAACGAGTGTGATGCTGCCACAGCAAGCCCGGATCAATCTCT	145
QY	100	GCAATGAGATGAGTGGCTTCAATTAAGTCAATCTGCGCATTTGTTCTCGAATGG	159
DB	146	CCAATTGGAATTTGTCGGCTGTAAATTTTTCATCTTTCCTTCGAAATGG	205
QY	160	GCTGAGTCAAGCGCTGCTTCAAAATCAATCTCAATTTCTGTAAACCTTCTCTA	219
DB	206	GCAATGTAATGAGATTTCCACAAAGTCATCTTTCATTTCAATTAAGAGCAGCAATG	265
QY	220	CCCTGAGATGATCTCTATACAGATGAGATGATGATGCTGCTGCTGAGCAAGATGGT	279
DB	266	CCCAAGATGATTAATCTTATACAGGATGACATGAAATGAATGCTTACTTGAAGAAATTAAGT	325

DB 358 CTAGTGCTCCCTCATGCTNNACACTGTGNGAAG 393

RESULT 11

US-09-878-574-1714

Sequence 1714, Application US/09878574

Patent No. US20020110548A1

GENERAL INFORMATION:

APPLICANT: Byrum, Joseph R.

APPLICANT: La Rosa, Thomas J.

APPLICANT: Thompson, Michael D.

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

TITLE OF INVENTION: Plants

FILE REFERENCE: 38-21(13401)B

CURRENT APPLICATION NUMBER: US/09/878,574

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 09/333,535

PRIOR FILING DATE: 1999-06-14

NUMBER OF SEQ ID NOS: 15775

SEQ ID NO 1714

LENGTH: 387

TYPE: DNA

ORGANISM: Glycine max

OTHER INFORMATION: Clone ID: LIB3028-034-Q1-B1-C4

US-09-878-574-1714

Query Match

Best Local Similarity 5.1%; Score 33.2; DB 10; Length 387;

Best Local Similarity 53.0%; Pred. No. 0.48;

Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 202 GTTGATTAACCTCTCTACCTGAGATGTGACTCTCTATACAGATGAGATGATGCT 261

DB 250 GCTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 309

QY 262 AGCTGCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 321

DB 310 AGATGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369

QY 322 ATGATCATTGCTCC 335

DB 370 AAGACAAAACTCC 383

RESULT 12

US-10-047-676A-3

Sequence 3, Application US/10047676A

Patent No. US20020123105A1

GENERAL INFORMATION:

APPLICANT: Qi, Fengxia

APPLICANT: Caulfield, Page W.

APPLICANT: Chen, Ping W.

TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS

FILE REFERENCE: UAB-17403/22

CURRENT APPLICATION NUMBER: US/10/047,676A

CURRENT FILING DATE: 2002-03-21

PRIOR APPLICATION NUMBER: US 09/627,376

PRIOR FILING DATE: 2000-07-28

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentin version 3.1

SEQ ID NO 3

LENGTH: 15567

TYPE: DNA

ORGANISM: Streptococcus mutans

US-10-047-676A-3

Query Match

Best Local Similarity 5.1%; Score 33.2; DB 12; Length 15567;

Best Local Similarity 53.3%; Pred. No. 3.8; Mismatches 78; Indels 6; Gaps 1;

Matches 96; Conservative 0;

QY 291 TCATATGAGCTCAGACGCTGGCTGATGATGATGATGATGATGATGATGATGATGAT 350

DB 6652 TCACATAGCTTAACTCGCTGGGAGACATACAGTGTGCTTACCTGCTACAGCAATAT 6711

QY 351 ATTAGCCAGATTGCTGGTGGTATGATATATATATATATATATATATATATATAT 410

DB 6712 AATTGAAAGATTGCTAATGATATGATATGATATGATATGATATGATATGATATG 6768

QY 411 GATTTATAGCAACCGTTGTTGTCACCGGCGATGATGATGATGATGATGATGATG 470

DB 6769 ---TTCTAGCAAGCCAGTTTAAATTATCTTGCATGATATATATATATATATG 6825

RESULT 13

US-09-974-300-5426

Sequence 5426, Application US/09974300

Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berka, Randy M.

APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods for Monitoring Multiple Gene

TITLE OF INVENTION: Expression

FILE REFERENCE: 10085,500-05

CURRENT APPLICATION NUMBER: US/09/974,300

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: 09/680,598

PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279,526

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 8481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5426

LENGTH: 1047

TYPE: DNA

ORGANISM: Bacillus clausii

US-09-974-300-5426

Query Match

Best Local Similarity 5.1%; Score 33; DB 10; Length 1047;

Best Local Similarity 47.1%; Pred. No. 0.97; Mismatches 150; Indels 6; Gaps 1;

Matches 139; Conservative 0;

QY 287 TTCTCATATGAGCTCAGACGCTGGCTGATGATGATGATGATGATGATGATGATG 346

DB 215 TTGCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 274

QY 347 ACACATTAGCCAGATGCTGCTGGTATGATGATGATGATGATGATGATGATGATGAT 406

DB 275 ATCAATTTGGAAATAGCGGACGATGCTGATGATGATGATGATGATGATGATGATG 334

QY 407 CARGGATTTAGCAACCGTTGTTGTCACCGGCGATGATGATGATGATGATGATGATG 466

DB 335 CAACG-----AAGCGCCAGTGCATTTAGCACCAGCAATGATGATGATGATGATG 388

QY 467 ATCCTTTCACAGAACGACCTTGTCTGTTGATGATGATGATGATGATGATGATGATG 526

DB 389 ACCGCTGTGACAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATG 448

QY 527 CCATTCAGAAAGAACTGCTGTGAGACATGATGATGATGATGATGATGATGATGATG 581

DB 449 CTGGCGGGGTTATCTTGTGCTGGGTTGGTGGCAAGGCCGATATGATGATGATGATG 503

RESULT 14

US-09-901-940-3

Sequence 3, Application US/09901940

Patent No. US20020150974A1

GENERAL INFORMATION:

APPLICANT: Hollaway, James L.

APPLICANT: Hefferman, Jane K.

APPLICANT: Taft, David W.

TITLE OF INVENTION: Placental Protein Having Multiple

TITLE OF INVENTION: EGF-like Domains

FILE REFERENCE: 99-43

CURRENT APPLICATION NUMBER: US/09/901,940

CURRENT FILING DATE: 2001-07-09

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3
LENGTH: 4527
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: This degenerate nucleotide sequence encodes the
OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
NAME/KEY: misc.feature
LOCATION: (1)...(4527)
OTHER INFORMATION: n = A,T,C or G
US-09-901-940-3

Query Match 5.1%; Score 33; DB 10; Length 4527;
Best Local Similarity 23.6%; Pred.No. 2.2;
Matches 73; Conservative 61; Mismatches 175; Indels 0; Gaps 0;

OY 11 CGAGCTCTCTCGGCTGTAACAAGATGAGACAGATGATGAGATGAGATGAG 70
DB 212 CMTGMSNMGNNTNNMNSNCARGATNARGARGARYTNACNARGACNAGCTNC 271
OY 71 TAACAAGGAGGCTGCTACTAGCTGCAAGTGAAGTGAAGTGAAGTGAAGTGA 130
DB 272 ARGCNTTTSNTTNTNGARYNGAYGCAATGATGATGATGATGATGATGATG 331
OY 131 GTAATCTGCGCATGTTTCTCAGATGAGTGAAGTGAAGTGAAGTGAAGTGA 190
DB 332 GYWSNCGMNSNCTNTTYYTNGARAAYCCNCAVGTNCARWSNCTNCGNATHGARWSNC 391
OY 191 CTCTCAATTTCTGTTAAACCTTCTTACCTCAGATGAGTGAAGTGAAGTGA 250
DB 392 ARGATTTTTCGCGNATGATGATGATGATGATGATGATGATGATGATGATG 451
OY 251 ATGAATGCTGAGTGAACAAGATGATGATGATGATGATGATGATGATGATG 310
DB 452 ARGATATATATYTNMGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 511
OY 311 GGGCTGATG 319
DB 512 TNGAYCARG 520

RESULT 15
US-09-983-965-5883/C
Sequence 5883, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengding
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983, 965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 5883
LENGTH: 398
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 63-LIB34-021-Q1-E1-H4
US-09-983-965-5883

Query Match 5.0%; Score 32.4; DB 10; Length 398;

Best Local Similarity 51.4%; Pred.No. 0.89;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

OY 116 CTTCATTAGTTCAGTATCTCCCATTTCTCAGAAATGGCTGCAAGTCAAGCCG 175

DB 229 CTTCAATTAAATTTCAAGTGTGATTTGATTTCTGTAATTTGCTGCTTTCTGTT 170
OY 176 TCGCTCAAAATCATCTCTCAATTTGTTGAATAAACCTTCTCTACCTGAGATGAGTCT 235
DB 169 TCAATTTGAAACTTCTCAGAGTTTGTGTTCTTCTTACCATACAGGTTAGTCTT 110
OY 236 TCTATACAGATGAGATGAATGTTCT 261
DB 109 TAATTAAAGAGAGACTCATTTTCT 84

Search completed: December 1, 2002, 21:57:41
Job time : 123 secs

QY	339	GTGGCAACACATTTACCAAGATTCGTGGCTATGTGTAAATTCATTTGACATGAT	398
Db	9062	AACGCTAACACATTTGCCAANAAGTACGTACAGATTTGGCGAACATGTAACAGTAC	9121
QY	399	AGTAAGACATGGGATTATAGCAACCGTTGTTTGTTCGACGGCGATGAACACTTGAT	458
Db	9122	AGCTCTAGCCCTACCAAGTATATATCCCAACATAATAGTCTCGTATGATACAAAT	9181
QY	459	GTGGAACAATCCCTTCACACAAGCGACCTGTCTGTGTGAACCTTGGAAATACCT	518
Db	9182	GTATGACCATCCAGTACTAGATATCTGAAAACATTAGAACTACGGCTATGA -GCT	9240
QY	519	AATTCCTCCATCAAGAAGAACTGGCGTGTGAGACTACGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGTCTCTAAGAAATCCCTACTAGCTGTGTGAGACCAAGCGAGGAGCTTTAGTGA	9300
QY	579	GCCTTCTCTGATTT	592
Db	9301	CCTGACCAATTAATTT	9314

RESULT 2

```

US-09-134-001C-1948
: Sequence 1948, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucellette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1948
: LENGTH: 1203
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1948

```

Query Match	8.48;	Score 55;	DB 4;	Length 1203;
Best Local Similarity	50.28;	Pred. No. 1.2e-08;		
Matches 167;	Conservative	0;	Mismatches 160;	Indels 6;
				Gaps 1

OY	272	AGATTGGGATCCCGCTTCATATCGAGCGCACAGCGCTGGGGCGATATGTATGATCATTTG	333
Db	194	AAAATCCCTGGAAGAGATATCAACATGTATCATTTAGGAGACTGGGGAGATCCGATTATATGCTG	255
OY	332	CTCCCTTCTGCTGTACACATATTAGCAGATTTCGGTGGGTGTATGTGATATATATTGGA	393
Db	254	CGCCAGCAACTGCTATATCTATCCGAAATTTAAGTGTGGATTCGTGATGATTTAAATTA	313
OY	392	CATGTATGTAGAAGCAATGAGGATTAATAGCAACCGTGTGTTGTTGCACCGCGCATGAACA	453
Db	314	CTTCTACTTACTTCTSC-----TACAAACAAACCAAAATTTGGTTGCAACCCGCAATGATG	366
OY	452	CTTTGATGTGAACACATCTTTTCACAGACGGCACCTTCTCTCTGATGAACCTTGAA	513
Db	368	TAAATATGTTATACATCTCCAGCTACTAATACATATATATGAAGTGTCTAATGATGAAGCGAT	428
OY	512	TCACCCTTAATTCCTCCCATCAAGAAGAACTGGCCCTGGAGACTACGGTATATGGCGAA	573
Db	428	ATTATTATTATGAACCTGGTAGTGCGCTTATTAGCATGTGTTATGTATGACAAAGGCGAA	488
OY	572	TGGCGAGCCCTTCTCTGATTTATTCACACTGTA	604
Db	488	TGGAAGAACCCATGCAAAATCTTATCTGTATTA	520

RESULT 3

```

US-09-134-001C-1939/c
; Sequence 1939, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1939
; LENGTH: 207
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1939

```

Query Match

	Query Match	6.7%	Score 44;	DB 4;	Length 207;	
	Best Local Similarity	54.3%;	Pred. No. 2.4e-05;			
	Matches 114;	Conservative 0;	Mismatches 90;	Indels 6;	Gaps	
QY	288 TCTTATATCGAGCTCAGACGCTGGGCTGATGTATGATCATTTGCTCTTGCTATAA	34				
Db	207 TCACATGATGATTCATTAGGAGACTGGGCGAGATTCGATTTATAGTCGCCAGCAACTGTAA	14				
QY	348 CACATTAGCCAGATTCGCTGGTGGGTTATGCATATCATTTGACATGTATATGAAGC	40				
Db	147 TACTATCGCAAAATTAAGTTGGTAGATTCGATGATATTAATTAATTCTTACATTACTTGC	88				
QY	408 ATGGGATTTATGCAACCGTGTGTTGTTGTCACCGCGATGAACACTTGGATGTGGAACA	46				
Db	87 -----TCAACACACCAAATTCGTTGGCACCCCAATGAAATGTAAATGTATTAACA	34				
QY	468 TCCITTTACAGACGGCACCTGTTCTGCT	497				
Db	33 TCCAGGTACTAATCAATATATGAAGTGT	4				

RESULT 4

US-08-822-774-18
; Sequence 18, Application US/08822774
; Patent No. 6183997

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

US-08-822-774-18

Query Match	6.6%	Score 43.4	DB 4	Length 1209
Best Local Similarity	51.3%	Pred. No. 0.0001		
Matches 101	Conservative 0	Mismatches 96	Indels 0	Gaps 0

OY	373	TTCATGATAATCTATTGACATGTATAGTAAGAACAATGGGATTTAATGCAACCCTGTTT	4322
Db	337	ATAGATGATACACTCCAGTAACACTACAGTGTCGACCAACACATTTCCCAATTTATG	3966
OY	433	GTTGCACCGGCGATGAACAACCTTTGATGTGGAAACAATCCTTTTACACAGAACGCCACCTTGTG	4322
Db	397	ATAACCCCCACCAATGCAATGAGACAAATGTACACGCGATCCCATAGTAAGAGGAGAACATTTGAA	4566
OY	493	TTCCTTGATGAACCTTG	509
Db	457	AGGTTAAAGAACTTGG	473

RESULT 5
US-09-632-711-18
; Sequence 18, Application US/09632711
Date: 4/11/2009 10:00:00 AM

GENERAL INFORMATION:
APPLICANT: HOGREE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:

Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington

```

?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: unknown
?      MOLECULE TYPE: DNA (genomic)
?      HYPOTHETICAL: NO
?      ANTI-SENSE: NO
?      SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-632-711-18

```

Query Match	6.6%	Score 43.4	DB 4	Length 1209
Best Local Similarity	51.3%	Pred. No. 0.0001		
Matches 101, Conservative	0	Mismatches 96	Indels 0	Gaps 0

[illegible]

RESULT 6
US-09-632-703B-18
; Sequence 18, Application US/09632703B

```

GENERAL INFORMATION:
APPLICANT: HOCREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF

```

Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:

STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.

ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-632-703B-18

Query Match 6.6%; Score 43.4; DB 4; Length 1209;
Best Local Similarity 51.3%; Pred. No. 0.0001;
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 313 GCTGATGTTATGATCATGCTCCCTTGTCTGCTAATAGCAATAGCAATGCTGTGG 372
DB 277 GCAGATTAAATTTGGTTGCTGCTGCCACTGCCACACAAATAGTAAGATTGCAATGGA 336
QY 373 TTATGTGATTAATCATATGACATGATATAGTAGAGCATGGATTTATAGCAACCGTTGTTT 432
DB 337 ATAGATGATCTCCAGTACATACAGTGTGACACAGCATTTCCCAATTCATTTATG 396
QY 433 GTTGACCCGGGAGACACATTTGATGTGGAACATCTTTCACAGAACGGCACCCTTGTG 492
DB 397 ATAGCCCGAGCATGACATGACATGTACAGGCATCCCATAGTAAGGAGAACATTTGAA 456
QY 493 TTGCTTGATGAACCTTGG 509
DB 457 AGTTAAAGAACCTTGG 473

RESULT 7

US-09-632-702-18
Sequence 18, Application US/09632702
Patent No. 6444428

GENERAL INFORMATION:

APPLICANT: HOGREFE, HOLLY
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
Extracts, PEP Protein Complexes, Isolated PEF Proteins,
and Methods for Purifying and Identifying Same

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: David J. Kulik, Evenson, McKown, Edwards &
Lenahan, P.L.L.C.

STREET: 1200 G Street, N.W. Suite 700

CITY: Washington

STATE: D.C.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/632,702

FILING DATE: 04-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/822,774

FILING DATE: 21-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kulik, David J.

REGISTRATION NUMBER: 36,576

REFERENCE/DOCKET NUMBER: 1486/43163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-8800

TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 1209 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-632-702-18

Query Match 6.6%; Score 43.4; DB 4; Length 1209;
Best Local Similarity 51.3%; Pred. No. 0.0001;
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 313 GCTGATGTTATGATCATGCTCCCTTGTCTGCTAATAGCAATAGCAATGCTGTGG 372
DB 277 GCAGATTAAATTTGGTTGCTGCTGCCACTGCCACACAAATAGTAAGATTGCAATGGA 336
QY 373 TTATGTGATTAATCATATGACATGATATAGTAGAGCATGGATTTATAGCAACCGTTGTTT 432
DB 337 ATAGATGATCTCCAGTACATACAGTGTGACACAGCATTTCCCAATTCATTTATG 396
QY 433 GTTGACCCGGGAGACACATTTGATGTGGAACATCTTTCACAGAACGGCACCCTTGTG 492
DB 397 ATAGCCCGAGCATGACATGACATGTACAGGCATCCCATAGTAAGGAGAACATTTGAA 456
QY 493 TTGCTTGATGAACCTTGG 509
DB 457 AGTTAAAGAACCTTGG 473

RESULT 8

US-09-221-017B-363
Sequence 363, Application US/09221017B
Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

INFORMATION FOR SEQ ID NO: 363:

SEQUENCE CHARACTERISTICS:

LENGTH: 4086 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO


```

NAME: Warbury, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/0055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2976 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-65

```

Query Match	5.3%	Score 34.8;	DB 3;	length 2976;
Best Local Similarity	50.0%;	Pred. NO. 0.14;		
Matches 144;	Conservative 0;	Mismatches 137;	Indels 7;	Gaps 2

Qy	310	TGGGCTATCTTTGAGTACATTCGCTCTTTGCTGGTAAACAATTGACCAAGTTCTGGT	369
Db	1755	TGGGCMATTCATCATTCATTGTTCACCTCGCAAGCGCAAAATACAAATTGCAAAATTGAGTGA	1698
Qy	370	GCGTTATGATTAATCTATTATGACATGTATATGTAAGACATGGGATTTATAGCAACCGTTG	429
Db	1695	GGTATTGCTATGATTGTAAGACATCAACGTGGTCTGTAACAGAG-----ACACCGAAA	1642
Qy	430	TTTGTTCACCGCGGAGTAAACACTTTGATGTGGAAACAATCCTTTACAGAAACGGACCTT	489
Db	1641	TTTATTTCGCTCTCTATGATGTGCAATGTGATGTATGAAAATTAACGTACCGACACAAATAT	1582
Qy	490	GTCCTTGCTGATTAACCTTGGAAATCACCTTAATTCCTCCCATTAAGAAGA- AACTGGCCTG	548
Db	1581	TAAATATTTTAAAGAAGATGGGATCATTTTATTCGAACCAAGGAAGGATTTCTTCGATG	1522
Qy	549	TGGAGACTACGGTAATGGCCCAATGCGTCAACCTTCTCGATTTATTC	596
Db	1521	TGGTATAGTTGCTTAAAGAGATGTGAAGAAACCGCTTCAAAATCGTTTC	1474

RESULT 13
 US-09-265-315-65/C
 : Sequence 65 Application US/09265315
 Patent No. 6187541
 :
 : GENERAL INFORMATION:
 :
 APPLICANT: Benton, Bret
 APPLICANT: lee, Yung J.
 APPLICANT: Malouin, Francois
 APPLICANT: Martin, Patrick K.
 APPLICANT: Schmid, Molly B.
 :
 APPLICANT: Sun, Dongxu
 :
 TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
 TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
 TITLE OF INVENTION: TARGET GENES
 NUMBER OF SEQUENCES: 111
 :
 CORRESPONDENCE ADDRESS:
 :
 ADDRESSEE: Lyon & Lyon
 STREET: 633 West Fifth Street
 STREET: Suite 4700
 City: Los Angeles
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 90071-2066
 :
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 MEDIUM TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: Word Perfect 5.1
 :
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/265,315
 FILING DATE: March 9, 1999
 CLASSIFICATION: 435

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Waldburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO.: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2976 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-65

```

Query Match	5.38;	Score 34.8;	DB 4;	Length 2976;
Best Local Similarity	50.08;	Pred. No. 0.14;		
Matches 144;	Conservative 0;	Mismatches 137;	Indels 7;	Gaps 2

QY	310	TGGCGCATGTTTAAAGATCATGCTCTCTTTGTCGGCAACACATTTGACCAAGATCTGCTG	369
Db	1755	TGGCGCATGCAATCATCTGTTTGGACCTGCCAAGGCAATATTCACAAATTTGAGTGTA	1696
QY	370	GCGTTATGATATATCATTTGACATGTATAGTAAGACATGGGATTTATAGCAAAACCGTTG	429
Db	1695	GGATTTGCTGATGATTTGGAGACATCAAGCTTGCTGCAACAGAG-----ACACGAAA	1642
QY	430	TTTGTTGCACCGGCGAGTGAACACTTTGATGTGGGAAACAATCCCTTTCACGAAGGACCTT	489
Db	1641	TTTATTTGCGCTGCTATGAATGTGCATATGATATGAATAAATTAACGTACCCACCAAAATAT	1587
QY	490	GTCCTTGCTGATGAACCTTGAATCACCTAAATTCCTCCATCAAGAGA- AACTGGCTG	548
Db	1581	TAAATATTTTAAAGAAGATGGTATCATTTTATTCGAACCAAGGAGGATTTTCTGACATG	1522
QY	549	TGCAAGCATACGGTAAATGGCCAAATGGCTAGCCTTCTCGATTTATTC	596
Db	1521	TGGTATAGTTGCTAAAGAGCGTATGGGAAGACCGCTTCAAACTGTTTC	1474

RESULT 14
 US-09-265-315-65/c
 Sequence 65, Application US/09265315
 Patent No. 6187541
 GENERAL INFORMATION:
 APPLICANT: Benton, Bret
 APPLICANT: Lee, Ying J.
 APPLICANT: Malouin, Francois
 APPLICANT: Martin, Patrick K.
 APPLICANT: Schmid, Molly B.
 APPLICANT: Sun, Dongxu
 TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
 TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
 TARGET GENES
 NUMBER OF SEQUENCES: 111
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Lyon & Lyon
 STREET: 633 West Fifth Street
 STREET: Suite 4700
 CITY: Los Angeles
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 90071-2066
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB

MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Walburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2976 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-65

Query Match 5.3%; Score 34.8; DB 4; Length 2976;
Best Local Similarity 50.0%; Pred. No. 0.14;
Matches 144; Conservative 0; Mismatches 137; Indels 7; Gaps 2;

QY 310 TGGCGTATGTTATGATGATGCTTCCTTGTCTGCTAACACATTAGCCAGATTGCTGTG 369
DB 1755 TGGCGMGATGCAATCATGTTGTCACCTGCAACGGCAATACAAATTCGAGTGTGA 1696
QY 370 GCGTTATGTATATCTTATGACATGTATAGTAAGACATGGATTAAGCAACCGTTG 429
DB 1695 GGTATGCTGATGATGTTGTCACATCAACGTTGCTAGCAACAG-ACACCGAAA 1642
QY 430 TTTGTTGACCGGGGATGACACTTGTGATGTAACAAATCCTTTCACAGAACGGACCTT 489
DB 1641 TTTATTTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1582
QY 490 GTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 548
DB 1581 TAAATTTTAAAGAGATGGGTATCATTTATGACACAGAGGAGGATTTCTAGCATG 1522
QY 549 TGGAGACTAGCGTAATGCGCAATGCGCTGAGCGCTTCTGATTTATTC 596
DB 1521 TGGTATGTTGCTTAAAGAGATGATGAGAGAACCGCTTCAAAATTCGTTTC 1474

RESULT 15
US-09-266-417-65/c
Sequence 65. Application US/09266417
Patent No. 6228588
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,417
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Walburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/248
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2976 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-266-417-65

Query Match 5.3%; Score 34.8; DB 4; Length 2976;
Best Local Similarity 50.0%; Pred. No. 0.14;
Matches 144; Conservative 0; Mismatches 137; Indels 7; Gaps 2;

QY 310 TGGCGTATGTTATGATGATGCTTCCTTGTCTGCTAACACATTAGCCAGATTGCTGTG 369
DB 1755 TGGCGMGATGCAATCATGTTGTCACCTGCAACGGCAATACAAATTCGAGTGTGA 1696
QY 370 GCGTTATGTATATCTTATGACATGTATAGTAAGACATGGATTAAGCAACCGTTG 429
DB 1695 GGTATGCTGATGATGTTGTCACATCAACGTTGCTAGCAACAGAG-ACACCGAAA 1642
QY 430 TTTGTTGACCGGGGATGACACTTGTGATGTAACAAATCCTTTCACAGAACGGACCTT 489
DB 1641 TTTATTTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1582
QY 490 GTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 548
DB 1581 TAAATTTTAAAGAGATGGGTATCATTTATGACACAGAGGAGGATTTCTAGCATG 1522
QY 549 TGGAGACTAGCGTAATGCGCAATGCGCTGAGCGCTTCTGATTTATTC 596
DB 1521 TGGTATGTTGCTTAAAGAGATGATGAGAGAACCGCTTCAAAATTCGTTTC 1474

Search completed: December 1, 2002, 19:19:27
Job time : 105 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 17:42:20 ; Search time 302 Seconds
(without alignments)
4869.387 Million cell updates/sec

Title: US-09-868-300-7

Perfect score: 653
Sequence: 1 gcaatgcgcagcagctcctt.....caaaagatggaaccagctg 653

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653	100.0	653	21	AAA51411
2	443.2	67.9	973	21	AAC35405
3	151.6	23.2	3348	24	ABU40569
4	128	19.6	500	21	AAC93745
5	121.2	18.6	903	22	AAS30612
6	121.2	18.6	903	22	AAS28151
7	120.6	18.5	784	22	AAH05519
8	120.6	18.5	1208	21	AAZ34608
9	120.6	18.5	1699	22	ABA08875

C	10	120.6	18.5	1699	22	AA160412	Human polynucleoti
	11	120.6	18.5	1748	22	AA158626	Human polynucleoti
	12	120.6	18.5	2211	22	AAH15923	Human cDNA sequenc
	13	81.2	12.4	543	24	ABN69125	Streptococcus poly
	14	76.6	11.7	540	24	ABN69124	Streptococcus poly
	15	63.6	9.7	9909	19	AAV52145	Streptococcus pneu
	16	61.4	9.4	349980	21	AAE21611	Neisseria meningit
	17	61.4	9.4	837096	21	AAAB1489	N. meningitidis pa
C	18	60.2	9.2	4503	20	AAAX13086	Enterococcus faeca
C	19	59.8	9.2	20844	21	AAAB1460	N. meningitidis pa
	20	59	9.0	676	21	AAA01916	Human colon cancer
	21	55	8.4	1200	22	AAH53591	S. epidermidis ope
	22	55	8.4	1203	22	ABN92485	Staphylococcus epi
	23	55	8.4	2937	22	AAH54809	S. epidermidis gen
C	24	55	8.4	3150	22	AAH54529	S. epidermidis gen
	25	52	8.0	11842	22	AAAS30674	DNA encoding novel
	26	52	8.0	11842	22	AAAS28753	Genomic sequence #
	27	50.8	7.8	5692	23	ABL19456	Drosophila melanog
	28	49.8	7.6	92407	22	AAE28349	Genomic fragment #
	29	48.2	7.4	1218	24	ABK73693	Bacillus lichenif
C	30	44.8	6.9	5137	18	AAV74528	Staphylococcus aur
C	31	44	6.7	207	24	ABN92476	Staphylococcus epi
	32	44	6.7	787	19	AAAX14137	H. pylori GHP0 319
	33	43.4	6.6	1209	19	AAV53859	Polymerase enhanci
	34	41.6	6.4	300	20	AAZ12612	Human gene express
C	35	41.4	6.3	349980	22	AAH41224	Pyrococcus abyssi
	36	40.4	6.2	1260	22	AAH66739	C glutamicum codin
	37	40.4	6.2	1389	22	AAE72065	Corynebacterium q1
C	38	40.4	6.2	349980	22	AAH68529	C glutamicum codin
	39	38.4	5.9	4445	24	ABA01096	B. subtilis
	40	37.6	5.8	8700	14	AAQ42541	B. subtilis
	41	37.6	5.8	8700	14	AAQ42541	B. subtilis
C	42	34.8	5.3	2976	21	AAZ25496	S. epidermidis plas
C	43	34.8	5.3	2976	21	AAZ25496	Essential Staphylo
C	44	34.8	5.3	2976	22	AAE08065	Staphylococcus aur
C	45	34.6	5.3	1231	23	AAE64225	Staphylococcus aur

ALIGNMENTS

RESULT 1	
AA51411	AA51411 standard; cDNA: 653 BP.
ID	AA51411
XX	
AC	AAA51411:
XX	
DF	26-SEP-2000 (first entry)
XX	
DE	A. thaliana Vb89 (HAL3) cDNA.
XX	
KW	Cyclin-dependent protein kinase; CDK: CDC2a; Vb89: HAL3;
KW	cell cycle; interacting protein; environmental stress; growth regulator;
KW	herbicide; nematode resistance; plant breeding; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FH	key
FT	location/qualifiers
FT	1..651
FT	/*tag= a
FT	/product= CDC2b-interacting-protein
FT	/partial
XX	
PN	W0200036124-A2.
XX	
PD	22-JUN-2000.
XX	
PE	17-DEC-1999; 99WO-EP10084.
XX	
PR	17-DEC-1998; 98EP-0124062.
XX	
PA	(CROP-) CROPDESIGN NV.
XX	

PI De Veylder L, Boudolf VKCK, Torres Acosta JA, Inze D;
 XX
 DR WP1: 2000-431601/37.
 DR P-PSDB: AAY96816.
 XX
 Nucleic acids encoding plant cell cycle interacting proteins, useful
 PT for regulating plant growth and in recombinant DNA protocols
 XX
 Claim 1; Page 125-126; 152pp; English.
 XX

The Vb8 clone encodes the Arabidopsis thaliana HAL3 homologue, a halotolerant gene isolated in *Saccharomyces cerevisiae*. The Vb9 clone interacts with *A. thaliana* CDC2b (a cyclin-dependent protein kinase (CDK)), but not with CDC2a in the two-hybrid system. CDC2a and CDC2b are the only CDK genes to have been characterized in detail in *Arabidopsis thaliana*. They were used as bait in a two-hybrid screening assay with a cDNA library of a plant cell suspension as prey. The plant cell cycle interacting proteins identified were designated LVP15, PHO80-like protein, Vb33, Vb89, VbDAMP and VbHSF. The nucleic acids, vectors comprising them, the proteins they express, antibodies that bind to them and/or inhibitors of their protein expression and/or activity may be used for modulating the cell cycle in an animal or plant cell division and/or growth, for influencing the activity of cell cycle proteins in a plant or animal cell, as positive or negative regulators of cell proliferation, for modifying the growth inhibition caused by environmental stress conditions (e.g. to improve growth of plants in normal or suboptimal nutrient conditions, especially phosphorus), for use in a screening method for inhibitors or activators of cell cycle protein, as growth regulators, herbicides and/or for inducing nematode resistance in plants. The DNA sequences and their regulatory sequences may be used as markers in plant or animal cell and tissue cultures or as a marker in marker-assisted plant breeding. The regulatory sequences may also be used for the expression of heterologous DNA sequences during a stage of the cell cycle.

SQ Sequence 653 BP; 180 A; 135 C; 151 G; 187 T; 0 other;

Query Match	100.0%	Score 653; DB 21;	Length 653;
Best Local Similarity	100.0%;	Pred. No. 8.5e-199;	
Matches 653; Conservative	0;	Mismatches 0;	Indels 0; Gaps 0.0;

OY	1	GAATTCGGCACGAGCTCTTCTTCTGCGCTGTACCAAGATAGAAAGAAAGATGAATATGAA	60
Db	1	GAATTCGGCACGAGCTCTTCTTCTGCGCTGTACCAAGATAGAAAGAAAGATGAATATGAA	60
OY	61	GTGATACAGTAACGAAGAACCCCTCGATCTTACAGTCGAAGTGAAGTGTGCTTCA	120
Db	61	GTGATACAGTAACGAAGAACCCCTCGATCTTACAGTCGAAGTGAAGTGTGCTTCA	120
OY	121	ATTAAATTCAGTAAATCTTCGCAATTTGTTCTCAGAAATGGGCTGAAGTCAAGCCGTCGT	180
Db	121	ATTAAATTCAGTAAATCTTCGCAATTTGTTCTCAGAAATGGGCTGAAGTCAAGCCGTCGT	180
OY	181	TCAAAATCATCTCAATTTGCTGTGATAAACCTCTACCTCGAATGAGACCTCTAT	240
Db	181	TCAAAATCATCTCAATTTGCTGTGATAAACCTCTACCTCGAATGAGACCTCTAT	240
OY	241	ACAGATGAAGATGAATGGTCTACCTGGAACAAGATTGGTGAATCCCGCTCTTCATATGAG	300
Db	241	ACAGATGAAGATGAATGGTCTACCTGGAACAAGATTGGTGAATCCCGCTCTTCATATGAG	300
OY	301	CTCAGAGCGTGGCGTGAATATATATCATATGCTCTTTGTCGTGAACATTTAGCCAA	360
Db	301	CTCAGAGCGTGGCGTGAATATATATCATATGCTCTTTGTCGTGAACATTTAGCCAA	360
OY	361	ATTTCGGTGGGTTATGATGAATCTATTGCATGTATGATTAAGAGATGGGATTTATAGC	420
Db	361	ATTTCGGTGGGTTATGATGAATCTATTGCATGTATGATTAAGAGATGGGATTTATAGC	420
OY	421	AAACCGTTTGTGTGACACGGCGGATGAACATTGTGATGTGGAAACAATCCCTTTCACAGAA	480
Db	421	AAACCGTTTGTGTGACACGGCGGATGAACATTGTGATGTGGAAACAATCCCTTTCACAGAA	480

QY	481	CGGACACCTTGTCTTGCTTGATGAACTTGGATTCACCTAATCTCTCCATTCAGAGAAA	540
Db	481	CGGACACCTTGTCTTGCTTGATGAACTTGGAAATCAACCTAATCTCTCCATTCAGAGAAA	540
QY	541	CTGGCCTTGGAGACTAAGGTAATGGGGCAATGGCTCTCTCGATTTATTCCACT	600
Db	541	CTGGCCTTGGAGACTAAGGTAATGGGGCAATGGCTCTCTCGATTTATTCCACT	600
QY	601	GTTACACGTTCTGGGATGACCAAGCTGCTAAACAAGAGATGGAACCACTTTG	653
Db	601	GTTACACGTTCTGGGATGACCAAGCTGCTAAACAAGAGATGGAACCACTTTG	653

RESULT 2
AAC35405
ID AAC35405 standard; DNA; 973 BP

AC	AAC35405;
XX	
DT	17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10075.

KW Hybridisation assay; genetic mapping; gene expression control;

KW metabolic pathway; promoter; termination sequence; ss

OS *Arabidopsis thaliana*.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.9%; Score 443.2; DB 21; Length 973;
Best Local Similarity 83.7%; Pred. No. 1.8e-131;
Matches 502; Conservative 0; Mismatches 98; Indels 0; Caps 0;

```
QY 53 ATATGAGATGATACAGTACAAAGAGCCGTATCTTACTACTGCAAGTGAAGTG 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 ACATGGAAGTAAATACACACCGAGAGAGCCCTCGTACTACTGCTGCAAGTGAACCG 190
QY 113 TGCGTTCAATTAAAGTCAATCTCTGCCATTTGTTCTCGAATGGGCTGAAGTCAAG 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 TCGGTGATCAAAATGCGCAATCTCTGCCATTTGTTCTCGAATGGGCTGAAGTCAAG 250
QY 173 CCGTCGCTCAAAATCATCTCTCATTTCTGTGATAAACCCTTCTACTGAGATGGA 232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 CCGTCGTTACCAAAATCATCTCTCATTTCTGTGATAAACCCTTCTACTGAGATGGA 310
QY 233 CTCTCTATACAGATGAAGATGAATGTCTACTGAGAACAGATTGATCCGCTTCTC 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 CTCTGTATACAGATGAAGATGAATGTCTACTGAGAACAGATTGATCCGCTTCTC 370
QY 293 ATATGAGCTCAGACGCTGGGCTGATGTTATGATCATTGCTCTCTGCTAAACAT 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 ACATGAGCTTACAGCTGGGCTGATGTTATGATCATTGCTCTCTGCTAAACAT 430
QY 353 TAGCCAAAGATTGCTGGGCTGATGTTATGATCATTGATGATGATGATGATGATG 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 TAGCCAAAGATTGCTGGGCTGATGTTATGATCATTGATGATGATGATGATGATG 490
QY 413 ATATAGCAAAACCGTTGTTGTCACCGGAGATGACACTTTGATGGAACAATCCTT 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 ACTATACCAAAACCGTTGTTGTCACCGGAGATGACACTTTGATGGAACAATCCTT 550
QY 473 TCACAGAAAGCGACCTTCTTCTGCTGATGAACCTTGGAATACACCTTAATCTCCATCA 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 TCACAGAAAGCGACCTTCTTCTGCTGATGAACCTTGGAATACACCTTAATCTCCATCA 610
QY 533 AGAAGAAACCTGGCCGTGGAGACTACGTAATGGCGCAATGGCTGAGCTTCTGATTT 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 AGAAGAAACCTGGCCGTGGAGACTACGTAATGGCGCAATGGCTGAGCTTCTGATTT 670
QY 593 ATTCACACTGTTAGACTGTTCTGGAGTCAACAAGCTGTAACAAGAGATGGAACAGTT 652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 ATTCACACTGTTAGACTGTTCTGGAGTCAACAAGCTGTAACAAGAGATGGAACAGTT 730

RESULT 3
ABL40569
ID ABL40569 standard; DNA; 3348 BP.
XX
AC ABL40569;
XX
DT 17-JUN-2002 (first entry)
XX
DE Canarypox virus (CapV) non-essential region nucleotide fragment #1.
XX
KW Canarypox virus; CapV; recombinant; vaccine; gene; ds.
XX
OS Canarypox virus.
XX
PN JP2002045184-A.
XX
PD 12-FEB-2002.
XX
PF 01-AUG-2000; 2000JP-0233097.
XX
PR 01-AUG-2000; 2000JP-0233097.
XX
PA (JABG ) NIPPON ZEON KK.
XX
WPI: 2002-299189/34.
XX
PT A DNA region nonessential to the growth of a virus and a recombinant
PS canariopox virus by using it -
XX
PS Claim 1; Page 8-9; 14pp; Japanese.
XX
```

CC The invention provides a DNA region nonessential to the growth of a
CC canarypox virus (CapV). A recombinant CapV, in which an exotic gene is
CC inserted to the above DNA region nonessential to the growth of a CapV,
CC can be used in an expression vector and a vaccine. The present sequence
CC represents a nucleotide fragment of the DNA region non-essential to the
CC growth of the canarypox virus.

Sequence 3348 BP; 1278 A; 442 C; 521 G; 1107 T; 0 other;

Query Match 23.2%; Score 151.6; DB 24; Length 3348;
Best Local Similarity 62.8%; Pred. No. 9.5e-38;
Matches 235; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
QY 233 CTCTCTATACAGATGAAGATGAATGTCTACTGAGAACAGATTGATCCGCTTCTC 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 CTATTTATACGATTAAGATGATGATGATGATGATGATGATGATGATGATGATG 348
QY 293 ATATGAGCTCAGACGCTGGGCTGATGTTATGATCATTGCTCTCTGCTAAACAT 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 ATATGAGCTTACAGCTGGGCTGATGTTATGATCATTGCTCTCTGCTAAACAT 408
QY 353 TAGCCAAAGATTGCTGGGCTGATGTTATGATCATTGATGATGATGATGATGATG 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 TAGCCAAAGATTGCTGGGCTGATGTTATGATCATTGATGATGATGATGATGATG 468
QY 413 ATATAGCAAAACCGTTGTTGTCACCGGAGATGACACTTTGATGGAACAATCCTT 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 ATATTAATAAACCTTACTTCTTCTGCTGATGAACCTTGAATGATGATGATGATG 528
QY 473 TCACAGAAAGCGACCTTCTTCTGCTGATGAACCTTGGAATACACCTTAATCTCCATCA 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 TTACTGAGACACCATTAATAAAGCTTGAACATATGAGGCTGATGATGATGATG 588
QY 533 AGAAGAAACCTGGCCGTGGAGACTACGTAATGGCGCAATGGCTGAGCTTCTGATTT 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 AGAAGAAACCTGGCCGTGGAGACTACGTAATGGCGCAATGGCTGAGCTTCTGATTT 648
QY 593 ATTCACACTGTTAGA 606
    ||||| |||||
Db 649 CACGTGTTGTGAGA 662

RESULT 4
AAC93745/C
ID AAC93745 standard; cDNA; 500 BP.
XX
AC AAC93745;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:240.
XX
KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.
XX
OS Clenoccephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX
PA (HESK-) HESKA CORP.
XX
WPI: 2000-656323/63.
XX
PT Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX
PS Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX
```



```
DB 380 TCTGCACATTCGACCTGCGAGGCGGACACCCCTGCTGAGTCTCTCTTGATGCCAA 439
QY 348 CACATTAAGCCAGATTCGTGGTGCGTTATGTATATCATTCATGACATGATAGTAGAGC 407
DB 440 CACTCTGGGAGAGGTGGCAGTGCGCATCTGTGACACTTGTCTACTCGGTACGGGGC 499
QY 408 ATGGAGTTATAGCAACCCCTTGTGTGTCACCGCGCATGCAACACTTGTATGGAACAA 467
DB 500 CTGGACGCCGAGAACCCCTGCTCTTCTGCGCGGCATGAAACCGCCATGTGGAGCA 559
QY 468 TCCTTTCACAGAGGCGACCTTGTCTTCTGTATGACTTGGATACCCCTAATTCCTCC 527
DB 560 CCCGATCACAGGCGCACAGGTACACCAGCTCAAGGCGCTTGGTATGTGAGATCCCTG 619
QY 528 CATCAAGAGAAATGCGCCCTGCGAGACTACGCTAATGCGGCAATGCGTGA 578
DB 620 TGTGGCCAGAAAGCTGTGTGTGGAGATGAAGTCTCGGGGCCATGGCTGA 670
```

RESULT 6

AAS28151

ID AAS28151 standard: cDNA; 903 BP.

XX AAS28151;

DT 07-NOV-2001 (first entry)

XX Novel cDNA encoding for human respiratory antigen #283.

XX Human; respiratory antigen; respiratory disorder; throat disorder;
XX lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
XX anti allergic; anti asthmatic; anti inflammatory; olfactory;
XX respiratory active; ss.

OS Homo sapiens.

XX WO200155448-A1.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01333.

XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-020515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234023.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GH04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
XX
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettein H;
XX
XX WPI; 2002-352536/38.
DR P-PSDB; ABP28494.
XX
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX
PS Claim 7; Page 3778-3779; 4525pp; English.
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 543 BP; 178 A; 112 C; 84 G; 169 T; 0 other:
Query Match 12.4%; Score 81.2; DB 24; Length 543;
Best Local Similarity 53.1%; Pred. No. 1.4e-15;
Matches 173; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 292 CATATGACGCTGACACCGCTGGCTGATGATCATTCCTCTTTGTCGTAACACA 351
DB 214 CATATGCAATTTAGCAAAACCCACAGATTTATTTATTTGGACCTGCTGCTAAACT 273
QY 352 TTGGCCAAAGTTGGTGGTATGATGATATCTATTGACATGATAGTAAGACATGG 411
DB 274 ATTGCTCATCTAGCATATGATGATTCGTAATATTTGTAACAAGTGTTCCTTGCCCTA 333
QY 412 GATTATGACAAACCGTTGTTGTCACCGCGCATGGAACCTTTGATGGAGACATCT 471
DB 334 CCAAGCTACGACACCAAAATTAATAGCCCGACCATGAAATACCAAAATGTATCAAAACCT 393
QY 472 TTGACAGAACGCACTTGTCTGCTTGATGATGACTTGGAATCCCTAATTCCTCCATC 531
DB 394 ATTACTCAAGAAATATTAATTAACGCTTATCGACATAGAGGTTTACAGAAATTCCTCAAAA 453
QY 532 AAGAAGAACTGGCGCTGGAGACTAGCGTAATGGCGCAATGGGTGAGCCCTTCTGATT 591
DB 454 TCTAGCTTATTAGCTGTGAGATTAAGACACTGCTGCTTGGCTGATATTGATGATTATC 513

QY 592 TATTCACACTGTAGACTGTTCTGCGA 617
DB 514 TTAGCTACTATTGATACGATTTGGA 539
RESULT 14
ID ABN69124 standard; DNA; 540 BP.
XX
XX ABN69124;
AC
XX
XX
DT 01-JUL-2002 (first entry)
XX
XX Streptococcus polynucleotide SEQ ID NO 6161.
DE
XX Streptococcus GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus agalactiae.
OS
XX
XX WO200234771-A2.
PN
XX
XX
PD 02-MAY-2002.
XX
XX
PF 29-OCT-2001; 2001WO-GH04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettein H;
XX
XX WPI; 2002-352536/38.
DR P-PSDB; ABP28493.
XX
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX
PS Claim 7; Page 3778; 4525pp; English.
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 540 BP; 184 A; 115 C; 84 G; 157 T; 0 other:
Query Match 11.7%; Score 76.6; DB 24; Length 540;
Best Local Similarity 53.5%; Pred. No. 4.3e-14;
Matches 160; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 292 CATATGACGCTGACACCGCTGGCTGATGATCATTCCTCTTTGTCGTAACACA 351
DB 214 CATATGCAATTTAGCAAAACCCACAGATTTATTTATTTGGACCTGCTGCTAAACT 273

OY	352	TTACCCAAAGATTGCTGGTGGTATTGGAATAATCTATTGACATGTAAGTAACAGCATGG	411
Db	274	ATTCTCTACTCTAGCATATATGCTTCCTGGCTGAACATCGTAACAGCCGTCTTCCTGATATG	333
OY	412	GATTATATAGCAACCGCTTGTGTGGTGCACCGCGATGAACACTTGTGATGTGGCAACAATCCT	471
Db	334	CCCGATGGAACCTCCAAAACCTTATTCGACACACGACATGTAATCAAAAGATATCTCATATATCA	393
OY	472	TTTCACAGACGGCGACCTTGTCTTGCCTTGATGAACCTGGAATCAACCCATAATTCCTCCCATC	531
Db	394	ATCACCCACAGCTAATCATTTGATATTTTGAAAAAATATAGTTATATCAGAAATTTGAACCTCGT	453
OY	532	AAGAAGAAACAGCGCTCTGGAGACATACGGTAATATGGCCCATTTGGCTGTAGCCTTCTCTGAT	590
Db	454	ATTATGTTTGCCTCTGTGGTGAATACCGGTCAAGGACCCCTGCTGTATATCTACTAT	512

CC	XX	RESULT 15
CC	XX	AAV52145
CC	ID	AAV52145 standard; DNA; 9909 BP.
CC	AC	
CC	XX	AAV52145;
CC	DT	23-OCT-1998 (first entry)
CC	XX	
DE	XX	Streptococcus pneumoniae genome fragment SEQ ID NO:12.
CC	XX	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW	XX	computer readable medium; vaccine; pharmaceutical composition; ds.
OS	XX	Streptococcus pneumoniae.
XX	XX	
PN	XX	MO9818931-A2.
XX	PD	
XX	PD	07-MAY-1998.
PF	XX	30-OCT-1997; 97WO-US19588.
PR	XX	31-OCT-1996; 96US-0029960.
PA	XX	(HUMA-) HUMAN GENOME SCI INC.
PI	XX	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI	XX	Kunsch CA, Rosen CA;
DR	XX	WPI; 1998-272225/24.
PT	XX	
PT	XX	Computer-readable medium with recorded Streptococcus pneumoniae
PT	XX	polynucleotide sequences - useful in diagnostic kits and assays, and
PT	XX	pharmaceutical compositions and vaccines for Streptococcus
PS	XX	pneumoniae
PS	XX	
PS	XX	Claim 1; Page 225-230; 1409p; English.
CC	XX	
CC	XX	The present invention describes a computer readable medium which has
CC	XX	the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC	XX	recorded on it, or a representative fragment or a sequence at least 95%
CC	XX	identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC	XX	SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC	XX	Streptococcus pneumoniae. The present invention also describes an
CC	XX	isolated nucleic acid molecule encoding a homologue of any of the
CC	XX	fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC	XX	nucleic acid molecule is produced by a process comprising: (a) screening
CC	XX	a genomic DNA library using as a probe a target sequence defined by any
CC	XX	of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC	XX	library which contain sequences that hybridise to the target sequence and
CC	XX	isolating the nucleic acid molecules from the members; or (b) isolating
CC	XX	mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC	XX	molecules whose nucleotide sequence is homologous to amplification
CC	XX	primers derived from the fragment of the S. pneumoniae genome to prime
CC	XX	the amplification and isolating the amplified sequences. The computer
CC	XX	readable medium can be used in a computer-based system for identifying
CC	XX	

CC fragments of the *S. pneumoniae* genome of commercial importance, or
CC expression modulating fragments of the *S. pneumoniae* genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for *S. pneumoniae*.
XX
SQ Sequence 9909 BP; 2919 A; 2322 C; 1713 G; 3044 T; 1 other;

expression modulating fragments of the *S. pneumoniae* genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for *S. pneumoniae*.

CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for *S. pneumoniae*.
CC

CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for *S. pneumoniae*.
CC

SQ Sequence 9909 BP; 2919 A; 2232 C; 1713 G; 3044 T; 1 other;

Query match 9.78; Score 63.6; DB 19; Length 9909;

Best Local Similarity 52.28; Pred. No. 2.6e-09;
Matches 164; Conservative 0; Mismatches 149; Indels 1; Gaps 1;

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTAATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCAAATCATATATGCAACCTTTGGAAAAAAGAGATTTTATTTATATTCGTGGTACCTGC	9061
Qy	339	GTCGTCTAACACATTTAGCCAAAGATTGCTGGTGGTATGTGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACATTTATGCAAAACTAGCTACGATTTTGGGGCAACATGTATACCAGTAC	9121
Qy	399	AGTAAGACATGGGATTTATAGCAAAACCGTTGTTTGTTCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATATGAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGCTTGTGTATGAACTTGAATCACCT	518
Db	9182	GTATAGCACATTCACGATTAATCTGAAATAACATTAGCAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCTGTGTGAGACTATAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGTGAGACCGAGACAGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTAATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCAAATCATATATGCAACCTTTGGAAAAAAGAGATTTTATTTATATTCGTGGTACCTGC	9061
Qy	339	GTCGTCTAACACATTTAGCCAAAGATTGCTGGTGGTATGTGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACATTTATGCAAAACTAGCTACGATTTTGGGGCAACATGTATACCAGTAC	9121
Qy	399	AGTAAGACATGGGATTTATAGCAAAACCGTTGTTTGTTCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATATGAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGCTTGTGTATGAACTTGAATCACCT	518
Db	9182	GTATAGCACATTCACGATTAATCTGAAATAACATTAGCAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCTGTGTGAGACTATAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGTGAGACCGAGACAGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTAATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCAAATCATATATGCAACCTTTGGAAAAAAGAGATTTTATTTATATTCGTGGTACCTGC	9061
Qy	339	GTCGTCTAACACATTTAGCCAAAGATTGCTGGTGGTATGTGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACATTTATGCAAAACTAGCTACGATTTTGGGGCAACATGTATACCAGTAC	9121
Qy	399	AGTAAGACATGGGATTTATAGCAAAACCGTTGTTTGTTCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATATGAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGCTTGTGTATGAACTTGAATCACCT	518
Db	9182	GTATAGCACATTCACGATTAATCTGAAATAACATTAGCAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCTGTGTGAGACTATAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGTGAGACCGAGACAGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTAATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCAAATCATATATGCAACCTTTGGAAAAAAGAGATTTTATTTATATTCGTGGTACCTGC	9061
Qy	339	GTCGTCTAACACATTTAGCCAAAGATTGCTGGTGGTATGTGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACATTTATGCAAAACTAGCTACGATTTTGGGGCAACATGTATACCAGTAC	9121
Qy	399	AGTAAGACATGGGATTTATAGCAAAACCGTTGTTTGTTCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATATGAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGCTTGTGTATGAACTTGAATCACCT	518
Db	9182	GTATAGCACATTCACGATTAATCTGAAATAACATTAGCAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCTGTGTGAGACTATAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGTGAGACCGAGACAGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTAATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCAAATCATATATGCAACCTTTGGAAAAAAGAGATTTTATTTATATTCGTGGTACCTGC	9061
Qy	339	GTCGTCTAACACATTTAGCCAAAGATTGCTGGTGGTATGTGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACATTTATGCAAAACTAGCTACGATTTTGGGGCAACATGTATACCAGTAC	9121
Qy	399	AGTAAGACATGGGATTTATAGCAAAACCGTTGTTTGTTCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATATGAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGCTTGTGTATGAACTTGAATCACCT	518
Db	9182	GTATAGCACATTCACGATTAATCTGAAATAACATTAGCAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCTGTGTGAGACTATAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGTGAGACCGAGACAGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGTCTTGTGATGAACTTGAATCACCT	518
Db	9182	GTATGACCATTCACGATACCTCAGAAATTAATCTGAAAACATTAAGAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCCTGTGTGAGACTAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGAAATCCCTACTACCTTGTGTGAGACCGAGACGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGACATATCCTTTACACAGAACGGACCTTGTCTTGTATGTAACCTTGATCACCT	518
Db	9182	GTATAGACATTCACGATTAATCTGAAATAACATTAAGAACTACGGCTATCA - GCT	9240
Qy	519	AATTTCTCCCATCAGAGAAACTGGCCTGTGAGACTAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGGAGACACGAGACAGAGACTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGACATATCCTTTACACAGAACGGACCTTGTCTTGTATGTAACCTTGAATCACCT	518
Db	9182	GTATGACCATTCACAGTACTACGAAATTAATCTGAAAACATTAAGAACTACGGCTATCA - GCT	9240
Qy	519	AATTTCTCCCATCAGAGAAACTGGCTGTGTGAGACTAGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGGAATCCCTACTACCTTGTGTGAGACACGAGACAGAGACTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGTCTTGTGATGAACTTGAATCACCT	518
Db	9182	GTATGACCATTCACGATACCTCAGAAATTAATCTGAAAACATTAAGAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCCTGTGTGAGACTAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGAAATCCCTACTACCTTGTGTGAGACCGAGACGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGTCTTGTGATGAACTTGAATCACCT	518
Db	9182	GTATGACCATTCACGATACCTCAGAAATTAATCTGAAAACATTAAGAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCCTGTGTGAGACTAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGAAATCCCTACTACCTTGTGTGAGACCGAGACGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGTCTTGTGATGAACTTGAATCACCT	518
Db	9182	GTATGACCATTCACGATACCTCAGAAATTAATCTGAAAACATTAAGAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCCTGTGTGAGACTAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGAAATCCCTACTACCTTGTGTGAGACCGAGACGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Qy	279	TGATCCGCTCTTCATATATGAGCTGACAGCGTCGGCTGATGTTATGATCATTTGCTCCCTT	338
Db	9002	TGATCAGGTCATATCATATGCAACCTTTGGAAAAAGAGATTTTATTTATTCGTGGTACCTGC	9061
Qy	339	GTCGCTATACACATTTAGCCAGATTTGCTGGTGTTATGATTAATCATTTAGCATGTAT	398
Db	9062	AACCTGCTAACACTATTGCAAAACTAGCTACGATTTTGGGGCAACATGTACACAGTAC	9121
Qy	399	AGTAAAGCATATGGATATTTAGCAAAACCGTTGTTGTTGGCACCGCGGATGACACTTTGAT	458
Db	9122	AGCTCTAGCCCTACACAAGTATATTTGCCAAACTAAATAGCTCCTGCTATGAAATACAAAAT	9181
Qy	459	GTGAGAACATTCCTTTACAGAACGCGACCTTGTCTTGTGATGAACTTGAATCACCT	518
Db	9182	GTATGACCATTCACGATACCTCAGAAATTAATCTGAAAACATTAAGAACTACGGCTATCA -GCT	9240
Qy	519	AATTCCTCCCATCAGAGAAACTGGCCTGTGTGAGACTAGGTAATGGCGCAATGGCTGA	578
Db	9241	GATTGCTCTTAAGAAATCCCTACTACCTTGTGTGAGACCGAGACGAGAGCTTTAGCTGA	9300
Qy	579	GCCCTTCCTGATTT	592
Db	9301	CCTCACAATTATTT	9314

Search completed: December 1, 2002, 19:17:39
Job time : 313 secs

Job time : 313 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2002, 18:30:25 ; Search time 3143 Seconds

(without alignments)
6046.494 Million cell updates/sec

Title: US-09-868-300-7

Perfect score: 653
Sequence: 1 gaattcgagcagatcctcctt.....caaaagatggaaccagttg 653

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

1: gb_Da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653	100.0	653	6 AX027396	AX027396 Sequence
2	530.6	81.3	2522	8 AT080192	U80192 Arabidopsis
3	530.6	81.3	95295	8 AC020889	AC020889 Genomic s
4	316.2	48.4	1867	8 AF166262	AF166262 Arabidops
5	316.2	48.4	6588	8 AB026641	AB026641 Arabidops
6	167.6	25.7	288539	14 AF198100	AF198100 Fowlpox v
7	166.8	25.5	138906	2 AC003946	AC003946 Oryza sat
8	166.2	25.5	631	8 PAB271130	PAB271130 Picea abl
9	120.6	18.5	1167	9 AF182419	AF182419 Homo sapi
10	120.6	18.5	1240	9 BC014409	BC014409 Homo sapi
11	120.6	18.5	2211	9 AK027491	AK027491 Homo sapi
12	113.8	17.4	1074	3 AY102691	AY102691 Drosophila
13	100.8	15.4	12855	8 SPAC15E1	SPAC15E1 Drosophila
14	99.8	15.3	2206	8 SCYR072C	SCYR072C S. cerevisia
15	99.8	15.3	4775	8 SCU01878	SCU01878 Saccharomyc
16	92.8	14.2	2187	8 CTHAL3GEN	CTHAL3GEN X88900 C. tropicali
17	92.8	14.2	155007	2 AC084047	AC084047 Trypanoso
18	85.2	13.0	3022	8 SCYOR053W	SCYOR053W S. cerevisia
19	85.2	13.0	54719	8 SCYXV55KB	SCYXV55KB Z74961 S. cerevisia
20	83.6	12.8	22239	3 CEP25H9	CEP25H9 Z81069 Caenorhabdi
21	81.2	12.4	10324	1 AE006562	AE006562 Streptococ
22	80.6	12.3	2145	8 SCYRLO86W	SCYRLO86W Z28088 S. cerevisia
23	79.6	12.2	10324	1 AE010042	AE010042 Streptococ
24	79.6	12.2	50511	1 AE014154	AE014154 Streptococ
25	72.6	11.1	895	11 CNS06K5R	AL402469 77 end of
26	71.8	11.0	1678	10 BC004779	BC004779 Mus muscu
27	70.4	10.8	12975	1 AE006291	AE006291 Lactococc
28	69.2	10.6	10029	1 AE008484	AE008484 Streptococ
29	69.2	10.6	13002	1 AE007423	AE007423 Streptococ
30	69.2	10.6	77743	2 SPNEU1910	AL449332 Streptococ
31	67.2	10.3	437	1 SMU48885	U48885 Streptococc
32	64.6	9.9	210185	2 AC125321	AC125321 Mus muscu
33	63.6	9.7	9909	6 BD003692	BD003692 Polynucle
34	62.8	9.6	165158	3 AC009912	AC009912 Drosophila
35	62.8	9.6	166146	2 AC009732	AC009732 Drosophila
36	62.8	9.6	167447	3 AC009459	AC009459 Drosophila
37	62.8	9.6	198244	3 AC007803	AC007803 Drosophila
38	62.8	9.6	313634	3 AE003454	AE003454 Drosophila
39	61.4	9.4	13249	1 AE002516	AE002516 Neisseria
40	61.4	9.4	349980	6 AX044033	AX044033 Sequence
41	58.2	8.9	326301	1 NMA622491	AL162757 Neisseria
42	57.2	8.8	10049	1 AE007681	AE007681 Clostridi
43	56	8.6	10066	1 AE010581	AE010581 Fusobacte
44	55.6	8.5	347400	1 AP003591	AP003591 Nostoc sp
45	55.2	8.5	110000	2 LMFCHR32_26	Continuation (27 o

ALIGNMENTS

RESULT 1
AX027396
LOCUS AX027396 653 bp DNA
DEFINITION Sequence 7 from Patent WO0036124.
ACCESSION AX027396
VERSION AX027396.1 GI:10188367
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
REFERENCE
AUTHORS 1 (bases 1 to 653)
TITLE de Veylder, L., Inze, D., Boudolf, V.R. and Torres, A.J.
Novel cell cycle genes and uses thereof

QY	87	TATCTTACTGCAAGTGAAGTGGCTTCAATTAAGTCACTAATCTGCGATTG	146
Db	1353	TATCTTACTGCAAGTGAAGTGGCTTCAATTAAGTCACTAATCTGCGATTG	1412
QY	147	TTTCTGAGAAATGGGCTGAAGTGAAGCGGCTGCTCAAAATCATCTCCATTTTCGTTGA	206
Db	1413	TTTCTGAGAAATGGGCTGAAGTGAAGCGGCTGCTCAAAATCATCTCCATTTTCGTTGA	1472
QY	207	TAAACCTTCTTACTGCAAGTGAAGTGGCTTCAATTAAGTCACTAATCTGCGATTG	266
Db	1473	TAAACCTTCTTACTGCAAGTGAAGTGGCTTCAATTAAGTCACTAATCTGCGATTG	1532
QY	267	GAAACAAGTTGGTATCCCGCTTCTCATATGAGCTCAGACGCGGCTGATGTTATGAT	326
Db	1533	GAAACAAGTTGGTATCCCGCTTCTCATATGAGCTCAGACGCGGCTGATGTTATGAT	1592
QY	327	CATTGCTCCTTGTGCTGCTACACATTAGCA-----	358
Db	1593	CATTGCTCCTTGTGCTGCTACACATTAGCAAGGCTAATAATATAGAAACGTAAGC	1652
QY	359	-----AGATTGCT	366
Db	1653	TTTAGATCTGGAATTTTAGCATTTTGGCTTCATATTGTTTTTTTGTGTCGACATTTGCT	1712
QY	367	GGGGGTTATGTAATTAATCTATTAAGTGAAGCATGGGATTAAGCAACCG	426
Db	1713	GGGGGTTATGTAATTAATCTATTAAGTGAAGCATGGGATTAAGCAACCG	1772
QY	427	TTGTTTGTGACCGGGGATGAACACTTGTGATGAGCAATCTCTTACAGAACGCGAC	486
Db	1773	TTGTTTGTGACCGGGGATGAACACTTGTGATGAGCAATCTCTTACAGAACGCGAC	1832
QY	487	CTTGTCTTGTGATGAACCTTGAATCACCCTAATTCCTCCATCAACAGAACTGCGC	546
Db	1833	CTTGTCTTGTGATGAACCTTGAATCACCCTAATTCCTCCATCAACAGAACTGCGC	1892
QY	547	TGTGAGACTAAGTGAATGCGCAATGGCTGAGCCCTTCTGATTTATTCACCTGTTAGA	606
Db	1893	TGTGAGACTAAGTGAATGCGCAATGGCTGAGCCCTTCTGATTTATTCACCTGTTAGA	1952
QY	607	CTGTTCTGGGAGTCAACAAGCTCGTAACAAAGATGGAACCACTGTTG	653
Db	1953	CTGTTCTGGGAGTCAACAAGCTCGTAACAAAGATGGAACCACTGTTG	1999
RESULT 3			
AC020889/c	95295 bp	DNA	linear
LOCUS	AC020889	Genomic sequence for Arabidopsis thaliana BAC T1N15 from chromosome	
DEFINITION	1, complete sequence.		
ACCESSION	AC020889		
VERSION	AC020889.3	GI:6938141	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C., Shinn, P., Altai, H., Bel, O., Chin, C., Chiu, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharisky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Federspiel, N.A., Theologis, A. and Ecker, J.R. Genomic sequence for Arabidopsis thaliana BAC T1N15 from chromosome		
AUTHORS	1		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 95295)		
AUTHORS	Ecker, J.R.		
TITLE	Direct Submission		

JOURNAL	Submitted (12-JAN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE	3 (bases 1 to 95295)
AUTHORS	Ecker, J.R.
TITLE	Direct Submission
JOURNAL	Submitted (08-FEB-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE	4 (bases 1 to 95295)
AUTHORS	Chen, R., Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Kim, C., Shinn, P., Altai, H., Bel, O., Chin, C., Chiu, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharisky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.
TITLE	Direct Submission
JOURNAL	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT	On Feb 8, 2000 this sequence version replaced gi:6939011.
FEATURES	
SOURCE	1. 95295 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="T1N15" complement(join(<1. .364,461. .506)) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="T1N15.1" /protein_id="AA079719.1" /db_xref="GI:8778711" /translation="MDASGSPFLPQLPSLDSLEPQVQVYVPHNDTDTITSVESILSFQME FGRSPSLSLCDLVGATLIDREFQNKQVYIPNDTDTITSVESILSFQME VSTCDMKNEDLVSQVLTANLEKDTAKHLQREQTLQEA" complement(join(1177. .1388,1493. .1524,1626. .1720, 1799. .1872,1960. .2104,2188. .2251,2337. .2399,2463. .2641, 2759. .3018,3107. .3269,3358. .3480,3831. .3902,4001. .4108, 4577. .4693,4765. .4903,4994. .5100,5199. .5303,5387. .5541, 5603. .5776,5832. .6061,6149. .6792,7208. .7313)) /note="similar to zwille protein emb CA011429.1: similar to ESTs gb AA7207232.1, dbj AA442417.1, gb A1993691.1, dbj AA41187.1, and gb AA651494.1" /codon_start=1 /evidence=not_experimental /product="T1N15.2" /protein_id="AA079718.1" /db_xref="GI:8778710" /translation="MAALVFVLIHLCVRSAYELDFSTRIGSASFMCYGIWRRK TDAPSEGGSGSREAGPVASGGRGSGRGGGOGGOGGGRYTPQPOGGGGRG GPPPOOOOGCPQOEYGRGRCGPRHGGCGTGGCGGSPGPPOROSVPELHQAT SPYTAQVSSQPTLSEVSPVPEPVYLAQGFEDLSVEQAPSAIDIPSSKAFNRP MRPGKGGSKRCIVKANHFAPLBDKLDHYDTITPEVYSKGNRAVMQALDNFND SHLGRPLPAYDGRKSLITAGPLPNSKEPFINLIDEEVAGGGRSSYCFRLRERKY VILVAADLHNLGMFLGKSDAPALQVLDIVLREPTSPWILLIIFRYIPG RSFYSPDKKQSLDGLSEWRGFPYSIRPTQMSLNGVMAKSYSTAFANPVI QPVCDLNRDISRPLSDADRVKIKKLRGVKVEYTRGMRKRYISGLTANATREL TFPVDERNTOKSVEYFHFETRYGRIHTQPLQDVGNSNPNLPEVCTIVGORS KLNRERITALLAVTCORPIDREKDLIQPOLNDYAKDNVAQEGIKISTLSAVAR ILPPNPKLYESREGTCPLPOVGMNMNMKNITNGVNMNICTINSROYODNARFE CLOLAQWYGVGAAPNEPLPVASRBEVEVLELTRYDADTSKLSQGEIDLIVY LPPNNSISYDGLKRICETEGIVSOCIIRKVRKMSKOWANALKINRYVGGARNVL VDLSKRITPLVSRPITIRGADYTHHPGDDSPSTAAYVASDWEPIRTKATLVCQ AHRQELDILFKEMKDPQGVNTGMIKTYVMILFNIFVIGELLIRSRSGHPLRI FYRQVSEGFQVLLYLEDAIKRACASLEAGQCPVPTFVYVGLKRRHTTFLAONHNR HSDVLSGNILIPETSTVDSKICHPEFEDFLCSHAAGIOTSRPAHYHVLMDGNFT DGLSLTNLNCYTTARCTRSVSIIVPAYVAHLAFAFRFYMERTSDSGSMASGAR CGGMAGRSTPGPVNAVRPLPALKEWRYKMYC" complement(join(11659. .11863,11978. .12075,12232. .12384, 12425. .12482,12570. .12662,12789. .12880,12982. .13092,

```
13176..13436,13522..13695))
/note="similar to 1-aminocyclopropane-1-carboxylate
deaminase (ACC) gb|AAD05069.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.3"
/protein_id="AAF79717.1"
/db_xref="GI:8778709"
/translation="MGRSLTISRYKLEIARSMGATSVPSNADELITTKKPYSPSAS
HLRLPSTHSLAHLPTPIHWNLPDLPGTSLWKRDQFTGMEISGNKVRLEPLMA
EVDVADHADVITIGIOSNHCRATATASNVLINSLILRTSKLADBDPGVGNLIV
ELVGAQNLHLSKEEYSSIGSEALNALKEKLEKGRKPYIPVGGNSLGTWYIEA
AETEEQNLNRPDDLKFDIVVACGGSTAGISGMGALKAKLTDSVKEPFYVO
VIAFVSCDDPQVYFVPGVGLDGIHAGVNSRIVNIHNAKGVAMNTSEELFVKV
ASTGYILDPVYSGRAITCLINETTKDPKCMGRKILFHTGSLGLYKIVQDMASLM
GMSRMDVSESVPRKDVGKMF"
/complement(join(17766..17891,18000..18095,18178..18290,
18375..18475,18551..18634,18739..18779,18849..18974,
19123..19196,19300..19595,19752..19807,19958..20041,
20151..20324,20522..20611,20677..20754,20864..20911,
21000..21112,21193..21256,21333..21406,21542..21581))
/note="similar to dihydroxyacetone kinase
sp|P4510|DHAK_CITFR; similar to ESTs gb|T42731.1,
gb|AA042761.1, and gb|N96033.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.4"
/protein_id="AAF79716.1"
/db_xref="GI:8778708"
/translation="MYPARKFTINPNVDVTEFIEGLVETPGIQLYLDGLPEIKVIR
ADVSANYNKVAIVISGSGHEPDAQYVEGMLTAIGDVPASPTVSIILRIAY
TGKGCGLVNTYNTGRLNGIAGAEQAKTEGPIEYTVIVGDDCALPPRGVARGG
LACTVYVHKYACAAAAGSLAEVAAAKASFMGCTMVALSTVCLPGVTSDRIGP
EKEKELGVHGEPGAADVIOPEVYVSHYLOOITSPETVYVITGNSVVMINIG
GTPLEMLTAAKCAVPRLOLEYGLADVRYTGSFMTSLDAGCCFCCVYESLITLCA
LILPFGSGSISIMKADOSTLERDLATNPASMPVTDIDLSRPAKIPVLPERS
TKREEESOGGRILEAAIEAAATVINLKMELMEMPQKVDGCGSGMCAAGAILD
MKRYPLMDAETVSEIGSSIRVMGSGSIIHPTVSEYTSYVNLCAAYAEKLANSS
ETTSKMSFALKSISISAVSKYGATAGTATGMDALIIPASVLEKELISGVDPIAAFLV
SAFAATYAGASTIHKRQAGRSSYSVAEIFASPDPGMAAAAMYSAAAAVAVEQGR
F"
join(21967..22138,22278..22378)
/note="unknown protein; similar to ESTs gb|AF162845.1,
gb|AI992553.1, and gb|AA042117.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.5"
/protein_id="AAF79693.1"
/db_xref="GI:8778685"
/translation="MALOWLILSVVAVEVITLVILPYPMLIKRNVOLVLSLIOP
AASIVAFAGROLDIYKRAEHLRISGSEVCTATERDRYEKSVTIP"
/complement(join(23638..24481,24564..24859,25106..25237))
/note="unknown protein; similar to EST gb|AI996045.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.6"
/protein_id="AAF79715.1"
/db_xref="GI:8778707"
/translation="MDAVANLFOSPSLAKQTSRLFTSPNCGFMRISGPOFRIR
ONSLSKFRPFGSGTTCVKSRSFVAKASGSDSTSIJLALQKSPVGGOLITLVS
HPILVPAVREDOLEQIDIDRAEODSADSVLETGDIIVLRIAEVEKEERLLEI
LYLVYQKFMADVNTLVPSITSSADSPGADVTDPTLDGELERASPEVEMIONHS
ETLSQAGROVERNEFYEAERTYAAVSSNDVGSFVGINASGFSDDMSOKRKYV
MSDEGETLQRAVTRRSREVGILIEKHTEALIGREVIYTPGQITDSSKDEHITKISFG
LKRILEAVTFGSLMDVESHVDRHYHFN"
join(26361..26450,26546..26588,26676..27014,27366..27706,
27823..28041)
/note="unknown protein; similar to ESTs emb|Z33851.1,
gb|T33009.1, and gb|AA394999.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.7"
13176..13436,13522..13695))
/note="similar to 1-aminocyclopropane-1-carboxylate
deaminase (ACC) gb|AAD05069.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.3"
/protein_id="AAF79717.1"
/db_xref="GI:8778709"
/translation="MGRSLTISRYKLEIARSMGATSVPSNADELITTKKPYSPSAS
HLRLPSTHSLAHLPTPIHWNLPDLPGTSLWKRDQFTGMEISGNKVRLEPLMA
EVDVADHADVITIGIOSNHCRATATASNVLINSLILRTSKLADBDPGVGNLIV
ELVGAQNLHLSKEEYSSIGSEALNALKEKLEKGRKPYIPVGGNSLGTWYIEA
AETEEQNLNRPDDLKFDIVVACGGSTAGISGMGALKAKLTDSVKEPFYVO
VIAFVSCDDPQVYFVPGVGLDGIHAGVNSRIVNIHNAKGVAMNTSEELFVKV
ASTGYILDPVYSGRAITCLINETTKDPKCMGRKILFHTGSLGLYKIVQDMASLM
GMSRMDVSESVPRKDVGKMF"
/complement(join(17766..17891,18000..18095,18178..18290,
18375..18475,18551..18634,18739..18779,18849..18974,
19123..19196,19300..19595,19752..19807,19958..20041,
20151..20324,20522..20611,20677..20754,20864..20911,
21000..21112,21193..21256,21333..21406,21542..21581))
/note="similar to dihydroxyacetone kinase
sp|P4510|DHAK_CITFR; similar to ESTs gb|T42731.1,
gb|AA042761.1, and gb|N96033.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.4"
/protein_id="AAF79716.1"
/db_xref="GI:8778708"
/translation="MYPARKFTINPNVDVTEFIEGLVETPGIQLYLDGLPEIKVIR
ADVSANYNKVAIVISGSGHEPDAQYVEGMLTAIGDVPASPTVSIILRIAY
TGKGCGLVNTYNTGRLNGIAGAEQAKTEGPIEYTVIVGDDCALPPRGVARGG
LACTVYVHKYACAAAAGSLAEVAAAKASFMGCTMVALSTVCLPGVTSDRIGP
EKEKELGVHGEPGAADVIOPEVYVSHYLOOITSPETVYVITGNSVVMINIG
GTPLEMLTAAKCAVPRLOLEYGLADVRYTGSFMTSLDAGCCFCCVYESLITLCA
LILPFGSGSISIMKADOSTLERDLATNPASMPVTDIDLSRPAKIPVLPERS
TKREEESOGGRILEAAIEAAATVINLKMELMEMPQKVDGCGSGMCAAGAILD
MKRYPLMDAETVSEIGSSIRVMGSGSIIHPTVSEYTSYVNLCAAYAEKLANSS
ETTSKMSFALKSISISAVSKYGATAGTATGMDALIIPASVLEKELISGVDPIAAFLV
SAFAATYAGASTIHKRQAGRSSYSVAEIFASPDPGMAAAAMYSAAAAVAVEQGR
F"
join(21967..22138,22278..22378)
/note="unknown protein; similar to ESTs gb|AF162845.1,
gb|AI992553.1, and gb|AA042117.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.5"
/protein_id="AAF79693.1"
/db_xref="GI:8778685"
/translation="MALOWLILSVVAVEVITLVILPYPMLIKRNVOLVLSLIOP
AASIVAFAGROLDIYKRAEHLRISGSEVCTATERDRYEKSVTIP"
/complement(join(23638..24481,24564..24859,25106..25237))
/note="unknown protein; similar to EST gb|AI996045.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.6"
/protein_id="AAF79715.1"
/db_xref="GI:8778707"
/translation="MDAVANLFOSPSLAKQTSRLFTSPNCGFMRISGPOFRIR
ONSLSKFRPFGSGTTCVKSRSFVAKASGSDSTSIJLALQKSPVGGOLITLVS
HPILVPAVREDOLEQIDIDRAEODSADSVLETGDIIVLRIAEVEKEERLLEI
LYLVYQKFMADVNTLVPSITSSADSPGADVTDPTLDGELERASPEVEMIONHS
ETLSQAGROVERNEFYEAERTYAAVSSNDVGSFVGINASGFSDDMSOKRKYV
MSDEGETLQRAVTRRSREVGILIEKHTEALIGREVIYTPGQITDSSKDEHITKISFG
LKRILEAVTFGSLMDVESHVDRHYHFN"
join(26361..26450,26546..26588,26676..27014,27366..27706,
27823..28041)
/note="unknown protein; similar to ESTs emb|Z33851.1,
gb|T33009.1, and gb|AA394999.1"
/codon_start=1
/evidence=not_experimental
/product="TIN15.7"
Query Match 81.3%; Score 530.6; DB 8; Length 95295;
Best local similarity 88.1%; Pred. No. 1,4e-149;
Matches 623; Conservative 0; Mismatches 4; Indels 80; Gaps 1;
CDS
27 CTGTAAACAAGTATAGAGAAAGATGATGGAAGTGGATACATACAGAAAGCCCTCG 86
|||
Db 84994 CTGATTTCAGTATAGAGAAAGATGATGGAAGTGGATACATACAGAAAGCCCTCG 84935
87 TATCTTACTAGCTCAAGTGGAGTGGCTGCTCAATTAAGTTCAGTAATCTCGCATTTG 146
TATCTTACTAGCTCAAGTGGAGTGGCTGCTCAATTAAGTTCAGTAATCTCGCATTTG
Db 84934 TATCTTACTAGCTCAAGTGGAGTGGCTGCTCAATTAAGTTCAGTAATCTCGCATTTG 84875
147 TTTCTCAAGATGGCTGAGTCAAAAGCCGTCGCTTCAAAATCATCTCATATTGCTGTA 206
TTTCTCAAGATGGCTGAGTCAAAAGCCGTCGCTTCAAAATCATCTCATATTGCTGTA
Db 84874 TTTCTCAAGATGGCTGAGTCAAAAGCCGTCGCTTCAAAATCATCTCATATTGCTGTA 84815
207 TTAACCTTCTACCTCAGATGAGTGTGACTCTCTATACAGATGAAGATGAATGCTTACCTG 266
TTAACCTTCTACCTCAGATGAGTGTGACTCTCTATACAGATGAAGATGAATGCTTACCTG
Db 84814 TTAACCTTCTACCTCAGATGAGTGTGACTCTCTATACAGATGAAGATGAATGCTTACCTG 84755
267 GAACAAGATTGGATCCGCTTCTCATATGAGCTCAGACGCTGGGCTGATTTATGAT 326
GAACAAGATTGGATCCGCTTCTCATATGAGCTCAGACGCTGGGCTGATTTATGAT
Db 84754 GAACAAGATTGGATCCGCTTCTCATATGAGCTCAGACGCTGGGCTGATTTATGAT 84695
327 CATTGCTCCTTTGCTGCTAACAATTAGCCA----- 358
|||||
Db 84694 CATTGCTCCTTTGCTGCTAACAATTAGCCAAGGACTAATAATCATAGAAACCTGAC 84635
359 -----AGATTGCT 366
|||||
Db 84634 TTTAGATCTGATTTTACCTATTGGCTTCATATTGTTTTTTTGTGCTGCAGATTGCT 84575
367 GGTGGTATGTGAATCTATTGACATGTATAGAGAGATGGATTAAGCAAAACCG 426
|||||
Db 84574 GGTGGTATGTGAATCTATTGACATGTATAGAGAGATGGATTAAGCAAAACCG 84515
427 TTGTTTGTGACCGCGCATGAACTTTGATGTGGAACATCTTTACAGCAAGCGGAC 486
|||||
Db 84514 TTGTTTGTGACCGCGCATGAACTTTGATGTGGAACATCTTTACAGCAAGCGGAC 84455
487 CTGTCCTGCTGATGATGAACTGGAATCACCCTAATTCCTCCATCAAGAAACTGGCC 546
|||||
Db 84454 CTGTCCTGCTGATGATGAACTGGAATCACCCTAATTCCTCCATCAAGAAACTGGCC 84395
```


QY	547	TGTGGAGACTACGGTAATGCGCATGCGTACGCTTCTGATTTATTCACACTGTAGA	606
Db	84394	TGTGGAGACTACGGTAATGCGCATGCGTACGCTTCTGATTTATTCACACTGTAGA	84335
QY	607	CTGTTCCTGGGAGTCAACAGCTCGTAAACAAAGATGGAACCACTTG	653
Db	84334	CTGTTCCTGGGAGTCAACAGCTCGTAAACAAAGATGGAACCACTTG	84288
RESULT 4			
LOCUS	AF166262	1867 bp	DNA linear
DEFINITION	Arabidopsis thaliana HAL3A protein (HAL3A) gene, complete cds.		PLN 04-MAY-2001
ACCESSION	AF166262		
VERSION	AF166262.1	GI:5802224	
KEYWORDS			
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1. (bases 1 to 1867)		
TITLE	Espinoza-Ruiz, A., Bellas, J. M., Serrano, R. and Culianez-Macia, F. A. Arabidopsis thaliana HAL3A: a flavoprotein related to salt and osmotic tolerance and plant growth		
JOURNAL	Plant J. 20 (5), 529-539 (1999)		
MEDLINE	20117504		
PUBMED	10652125		
REFERENCE	2. (bases 1 to 1867)		
AUTHORS	Culianez-Macia, F. A., Espinoza-Ruiz, A. and Serrano, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-JUL-1999) Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Camino de Vera s/n, Valencia 46022, Spain		
FEATURES	Location/Qualifiers		
source	1. 1867		
gene	/organism="Arabidopsis thaliana"		
gene	/db_xref="taxon:3702"		
gene	<954. .>1700		
gene	/gene="HAL3A"		
gene	join(<954. .1289,1407. .>1700)		
gene	/gene="HAL3A"		
gene	/product="HAL3A protein"		
gene	join(954. .1289,1407. .1700)		
gene	/gene="HAL3A"		
gene	/note="similar to Saccharomyces cerevisiae HAL3"		
gene	/codon_start=1		
gene	/product="HAL3A protein"		
gene	/protein_id="A051616.1"		
gene	/db_xref="GI:5802225"		
gene	/translation="MEMGRKRDROMENVTPRKRIVLAASGSVAALIKFGULCFIEWAEYRNVYTKSSLIHFLDKLSLPQEVITLYTDEDESSNNKIGDPLHLELRADVLVYAPLSANVLGK1AGGLCDNLTLCTIIRADYDKPLFVAPAMNTLMNNFETERNHLLSDELGITLLPIPIKRRACGAYGNGAMAEPSLIYSTVRLFEWSEAHQQTGTS"		
BASE COUNT	535 a 381 c 380 g 571 t		
ORIGIN			
Query Match	48.48;	Score 316.2;	DB 8;
Best Local Similarity	70.08;	Prod. No. 1.1e-84;	Length 1867;
Matches 502;	Conservative 0;	Mismatches 98;	Indels 117;
		Gaps 1;	
QY	53	ATATGGAAGTGTGATACAGTACAGGAAGCTCGATCTTACTAGTCCAGTGAAGTG	112
Db	982	ACATGGAAGTGTGATACAGTACAGGAAGCTCGATCTTACTAGTCCAGTGAAGTG	1041
QY	113	TGGCTCAATTAGTCTAGTATCTTCCACTTGTTCAGATGGGCTGAAGTCAAG	172
Db	1042	TGGCTCTATCAATATTCGGCAATCTTCCACTTGTTCAGATGGGCTGAAGTCAAG	1101
QY	173	CCGTCGGTCAAAATCAATCTCAATTTGGTGAATAAACCTTCTACTACAGATGTGA	232
Db	1102	CCGTCGGTCAAAATCAATCTTACATTTCTCGATAAACCTCTCTCTCCACAGAGTGA	1161

QY	233	CTCTCCTACAGATGAAGATGAATGTCCTTACTGGAAACAATTGGTGATCCCGTTCTTC	292
Dd	1162	CTCTGTATACGATGAAGATGAATGTCCTTACTGGAAACAATGGTGATCTCTGCCCTTC	1221
QY	293	ATATCGAGCTCAGACCGCTGGCGTATGTTATGATCAVTTGCCTCTTGTCTGTCAACACAT	352
Dd	1222	ACATCGAGCTTAGACGTTGGCGTATGTTTAGTCATTTGGCTCTTGTCTGTCAACACCT	1281
QY	353	TAGGCA-----	358
Dd	1282	TAGGCAAGGTATTTAACAAATCGACGAAAACCCTTAACCACTACTTACATTCAATTTGT	1341
QY	359	-----	358
Dd	1342	GTCAGCTGTGAAAAACATTCATGAGAGATTCATTCATCATATATTTTTTGTGACACTT	1401
QY	359	---AGATTGCTGGGGGTTATGTGATAATCTATTGACATGTATAGTAAGACATGGGATT	415
Dd	1402	TGTGATTTCTGGGGGCTTTGTATATCTTCTGACATTTGATATAGACCTGGGACT	1461
QY	416	ATACCAAACCCCTTTTGTTCGCACGGCGATGAACACTTTGATGNGAACAATCCTTTCA	475
Dd	1462	ATACCAAACCACTTTTGTTCCTTCACAGTATGAATACTTTGATGTGGAAACAATCCTTTCA	1521
QY	476	CAGAAGCGACCTTGTCTTGTCTTGATGAATCTTGAATCACCTATTTCTCCATCAAGA	535
Dd	1522	CTGAAAGGCATCTTTTGTCTCTTGATGAACGGAATCACACTTATTCCTCTCATCAAGA	1581
QY	536	AGAAACGCGCTGTGGAGACATCAGGTATAGCGGAATGGCGACGCTTCTGTGATTTAT	595
Dd	1582	AGAACTGTGCTGTGGAGACTACGTAATGAGAGGTATGCGCTGACGCCCTCTTTATCTATT	1641
QY	596	CCACTGTTAGACTTTCTTGGGAGTCCACAAAGCTCGCTTAAACAAGAATGGAACCAAGTT	652
Dd	1642	CCACTGTCAGACTCTTCTGGGAGTCTCAGGCTCATACGAACCGGTGAGACTGATT	1698
RESULT 5			
AB026641/c			
LOCUS	AB026641	6588 bp	DNA linear PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MBG14.		
ACCESSION	AB026641	BA000014	
VERSION	AB026641.1	GI:4757397	
KEYWORDS			
SOURCE			
ORGANISM	Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1 clone:MBG14.		
REFERENCE	Arabidopsis thaliana		
AUTHORS	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;		
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.		
	1 (stiles), Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.		
	kaneko,T.,		
	Structural analysis of Arabidopsis thaliana chromosome 3. II.		
	Sequence features of the 4,251,695 bp regions covered by 90 Pl, TAC		
	and BAC clones		
	DNA Res. 7 (3), 217-221 (2000)		
JOURNAL	20363099		
MEDLINE	2 (bases 1 to 6588)		
REFERENCE	kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research		
	Institute, Department of Plant Gene Research; 1532-3, Yana,		
	Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp,		
	Te1:81-438-52-3935, Fax:81-438-52-3934)		
COMMENT	Address for correspondence: kaos@kazusa.or.jp		
	For the latest information on annotation of this clone, please see		
	http://www.kazusa.or.jp/kaos/cgi-bin/seq.graph.cgi?c=mbg14		
	Genes with similarity to proteins in the databases are described in		
	'product' or 'note' qualifiers. Genes that have no significant		
	protein similarity are described as 'unknown protein'.		
	The software programs used to predict genes include: Grail		
	(Informatics Group, Oak Ridge National Laboratory,		
	http://compbio.ornl.gov/Graal-1.3//).		

gene GB:AF021350"
/codon_start=1
/product="ORF FPV001"
/protein_id="AAf44603.1"
/db_xref="GI:7271757"
/translation="MKMNPPIEYKKEKFLRDNSEYILPITCLTSVITSCLEA
ALFVAVDCKRDSFLEDTTATITSSSIATYDNIYIHCPRDMSHNGICLSTGK
VGRGQIQKCEKIGSDMIGKSEEMKALKINMGNDISRFVNDRAAASFDPNBECA
YGRSSVSEVPKVLTPSCSVRRYLVCNTDNTSYPTQSSFFNYE"
3367..4035
/gene="FPV002"
3367..4035
/gene="FPV002"
/note="identical to FPV259 Fowlpox GB:A06621"
/codon_start=1
/product="ORF FPV002 hypothetical protein"
/protein_id="AAf44597.1"
/db_xref="GI:7271751"
/translation="MIVEKIAAMLYPLCLRCGLNSVPRATCKVHCILLYPEVCC
ECMSFTIDSEHSCCTCCVPLILIRFRRVILPTLKATDCIRLPCVLIRPKRT
ICPLASWCRCFCPCPEVFLRCLFPCMLRRMRGRLTGVRGAFRDSRDPARGT
WVDWCEDDLVCWIMSPCCYVKRCIRTCDFTRKIFWYFIAPAGSPHPEPPLSRK
VFSS"
complement(4500..4871)
/gene="FPV003"
complement(4500..4871)
/gene="FPV003"
/note="identical to FPV258 Fowlpox PIR:C31685 (vaccinia
A40R), similar to Homo sapiens SW:Q07108"
/codon_start=1
/product="ORF FPV003 C-type lectin gene family protein"
/protein_id="AAf44606.1"
/db_xref="GI:7271760"
/translation="MKSLLIAMLMDCAINSNCRGPTSYNNKCIWNRDLKMHK
KTYSEASTCLIFPMTGLARSLIDNEKDMKFIKSGMGSLIRDDKKPEYKCAK
TDGKTGFSPCNMTYGFVCIID"
5125..5430
/gene="FPV004"
5125..5430
/gene="FPV004"
/note="identical to FPV257 Fowlpox GB:A06621"
/codon_start=1
/product="ORF FPV004 hypothetical protein"
/protein_id="AAf44598.1"
/db_xref="GI:7271752"
/translation="MNNDITFTLYCKNKKYVGGEGRRGKGTILLFHPINRVIGT
SAHQCKYTRIGIKLAVARHVSITIRCGSHAHKDRKSPFQKDFHICAGSA"
complement(5221..5589)
/gene="FPV005"
complement(5221..5589)
/gene="FPV005"
/note="identical to FPV256 Fowlpox PIR:E31685"
/codon_start=1
/product="ORF FPV005 ERG gene family protein"
/protein_id="AAf44604.1"
/db_xref="GI:7271758"
/translation="MALNLRVPEDESGIKVLAASDPLNYENDHTGDCFYARAR
NMEFCRARIYCAAGDTVKIYFLEGELIYSRVGSPADSGYVIRGNCVEPETS
SCFTLMCTSSYTYVYVME"
5931..7187
/gene="FPV006"
5931..7187
/gene="FPV006"
/note="identical to FPV255 Fowlpox SW:P14361 (vaccinia
C10L), similar to SW:P03296"
/codon_start=1
/product="ORF FPV006 C4L/C10L-like gene family protein"
/protein_id="AAf44589.1"
/db_xref="GI:7271753"
/translation="MEFSCTGTNNQQLAVHRTFTSFKEKLLVNLGISDLNDIKNI
CEGSKTFEPKRETLISIKDRKSQIYFENSLNDLKKHALIYDELSTVDVSVTVE
NPTLIMYEKGDFARHDRSVESKNIITGVHLLIYLOPEPGEYVIYIDNNYSVKL
KTHLEFKTIEHESITVESGRKCVALFDVLEKELASNTAVTGSIEYLGKINLYDRE

gene NDLOICDNIWERTEDKEYSLGIMSDRSGRCIKSHNNGSIYVRKEEYSPDALCI
YNNAEVDEIWTGDKKHIITWSTIDKTKTSTFIPIDPVYERIKAISSKEHEKDYDRF
CNSRTEYICCSVSKYYFDLPKTDLHEVINSIDYDRKSVGTPDWYTLPLEVKOTILG
NMSYEELFNIVRGIMALEEDNEVGC"
7599..7817
/gene="FPV007"
7599..7817
/gene="FPV007"
/note="identical to FPV254 Fowlpox GB:D00295"
/codon_start=1
/product="ORF FPV007 hypothetical protein"
/protein_id="AAf44600.1"
/db_xref="GI:7271754"
/translation="MYSVFTRIKYKIRKNKTEMIATYDRNAVYKKGISGILLRD
NGRNMILNIIYDFSSRMQWLVRSVL"
7681..8184
/gene="FPV008"
7681..8184
/gene="FPV008"
/note="identical to FPV253 SW:P14370 (vaccinia A40R),
similar to Gallus gallus GB:AB015628"
/codon_start=1
/product="ORF FPV008 C-type lectin gene family protein"
/protein_id="AAf44607.1"
/db_xref="GI:7271761"
/translation="MPTLLRKQVSESCVAITVIGLICLLIFTLVVVTKWYAPY
FSKVCPEDEWIGNSKCYFTINETNMDSKCLDVMDSLRIPENIETLNFVSYGSG
SYWIDIQNRKIPQINFSLYEGVNDICLLFDTSNIEMSCIFHERTICVKEDRYTH
WYEPYMR"
8048..8248
/gene="FPV009"
8048..8248
/gene="FPV009"
/note="identical to FPV252 Fowlpox GB:D00295"
/codon_start=1
/product="ORF FPV009 hypothetical protein"
/protein_id="AAf44601.1"
/db_xref="GI:7271755"
/translation="MKNALMIFVYLLRVTLSKCLVYFKERYVLKKIDYPIGIPNTC
VRFSTSLYSILYLVNRSKY"
complement(9123..10190)
/gene="FPV010"
complement(9123..10190)
/gene="FPV010"
/note="Fowlpox SW:P14369, similar to Homo sapiens
GB:AB006423"
/codon_start=1
/product="ORF FPV010 Serpin gene family protein"
/protein_id="AAf44354.1"
/db_xref="GI:7271508"
/translation="MGSILVRLKLELYVPGKDCISPGVYTIILMIMIGCKKETRDKI
KDLIGFNGVVPIDPKSEYVEVYDDKDELINSIMIEGYPDKRPFINSIDYNA
KYVSPDDTISPTINKKVELSTGLIKDESIADDIRLAINIVLYKSKWKPPDTE
LTSKHPKRYNGTDVMDIDVAVYTHDEDRSQVAVLEEDREFWFIIPDS
VTSIGDVSDLNKGNKINIKIISKDMTLKTIYVLPFELEDVDVLDKDALIHGCCNDL
FKSGELVIGSDFTKLIRIGNIRKSVIKVDEYGEAASVTESCITDGIKIPIVKANP
FMFLVAVOTKPIELFLGIFOG"
complement(10276..11112)
/gene="FPV011"
complement(10276..11112)
/gene="FPV011"
/note="similar to Bos taurus PIR:S32367"
/codon_start=1
/product="ORF FPV011 alpha-SNAP"
/protein_id="AAf44355.1"
/db_xref="GI:7271509"
/translation="MOKIKLVSPFEKKNLFGPPIETIEBAAGVSHAANLFAVOLMEC
AGKAFKSGDMLQKNKNSIAASSFVDANAARKIDSYBAINCLSKALEVYVCIDKF
YTVARCKIMNIAIYENDILELDRKAIHYENASGYVGGEGYNNKLSDDCMLIARLSIOK
EDDPDRAGKIEOVGYNNMMMLSKYESRHOLLAIKMCYLCSDVSRRAKRSIDKXKIDFP
AFKDFKCKRIEIKLIAACETKNETFTSAIEEDYDHGNTIDEALMSMLLTIRKATFDE
VE"
complement(12038..13033)

Accession	Sequence	Position
Db 113058	TCTTGAGACATCAACCTGTAGGTATATCTTTGGTCCCTCCCATTTACCAAAAGCGCTGGC	113117
QY 546	CTGTGGAGACTACGGTATGCGCGCAATGGCTGAGCCTTCTCTGATTTATTTCCACTGTTAG	605
Db 113118	CTGTGGTATATATGATATGGTCAGATGGCTGAGGCTTCTGTATCGATATTCACCGCTAG	113177
QY 606	ACTGTTCGAGATCACAGGCTCGTAAACAA	637
Db 113178	GCTTGCTTGCAAGACAGCCACTTAATACAA	113209

RESULT 8	
PAB271130	
LOCUS	PAB271130
DEFINITION	631 bp mRNA linear PLN 18-JAN-2002
	Picea abies partial mRNA for putative halotolerance protein HAL3
	homolog, (pPA0038 gene).
ACCESSION	AJ271130
VERSION	AJ271130.1
KEYWORDS	GI:12583570
SOURCE	halotolerance protein HAL3 homolog; pPA0038 gene.
ORGANISM	Norway spruce.
	Picea abies

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1	Schubert, R., Mueller-Starck, G. and Riegel, R.	Development of EST-PCR markers and monitoring their intrapopulational genetic variation in <i>Picea abies</i> (L.) Karst	Theor. Appl. Genet. 103 (8), 1223-1231 (2001)	2 (bases 1 to 611)

JOURNAL Submitted (24-JAN-2000) Dr. Schubert R., Faculty of Forest Sciences, Section of Forest Genetics, Technical University of Munich, Am Hochanger 13, D-85354 Freising, GERMANY

FEATURES .location/Qualifiers

FEATURES	location/qualifiers
source	1. .631

```

/cell_type="suspension cells"
/clone_lib="elictor-induced cdna library"
1..363
/gene="pPA0038"
<1..363
CDS

```

BASE COUNT ORIGIN	174 a	135 c	126 g	195 t	1 others
----------------------	-------	-------	-------	-------	----------

Query Match	25.5%	Score 166.2;	DB 8;	Length 631;
Best Local Similarity	68.9%	Pred. No. 3.4e-39;		
Matches 228;	Conservative	0;	Mismatches 103;	Indels 0;
				Gaps

280 GATCCGTTCTTCATATCGAGCTCAGACCGTGGGCTGATGTTAATGATCATTTGCTCCTTG 33

D6 10 GGTACTGTACTGCATATAGACTCCGGCATTGGCGTGATGCTATCGTGTGATTTGCTCCACTA 69

340 1C10C1AACACAI1AACCAAGAA11GC1GG1GGG1TA1G1GTA1AAC1TA1T1GACAI1ATA 35

400 CTGACAGCAATCCGAATTAACTCCAAAACCCTCCTTTCCTTTCTCATCCCCTCACACCTCTGC
A

Db 130 ATACGTCATGGGACCTTTAAACAAGCCTCTCTTTGTAGCTCCMGATGAATACCTTCATG 18

[illegible]

RESULT	9			PRI	20-SEP-2000
LOCUS	AF182419	1167 bp	mRNA	linear	
DEFINITION	Homo sapiens MDS018 (MDS018)	mrna,	complete cds.		
ACCESSION	AF182419				
VERSION	AF182419.1	GI:10197637			
KEYWORDS	.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 1167)	Huang, C., Qian, B., Tu, Y., Gu, W., Wang, Y., Han, Z., and Chen, Z.	Novel genes expressed in hematopoietic stem/progenitor cells from myelodysplastic Syndromes patient

REFERENCE
2 (bases 1 to 1167)
AUTHORS Huang, C., Qian, B., Tu, Y., Gu, W., Wang, Y., Han, Z. and Chen, Z.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1999) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, People's Republic of China

FEATURES	Location/Qualifiers
source	1. .1167

```

gene      /note="isolated from a Myelodysplastic Syndrome patient"
          1..1167
          /gene="MDS018"
CDS       252..635
          /gene="MDS018"

```

BASE COUNT	274 a	318 c	338 g	237 t
------------	-------	-------	-------	-------

Query Match 18.58; Score 120.6; DB 9; Length 1167;

	Matches	207;	Conservative	0;	Mismatches	144;	Indels	0;	Gaps	0;
OY	228	TGTGACTCTCTATACAGATGAGATGTCCTAGCTGGAACAACAGATTGGTCGATCCCGT	287							

Db 218 TGTACCCCTCTACAGCGACGCTGATGATGGAGATGTGGAAGGCCGCTCTGACCCAGT 277

288 TCTTCATATCGAGCTCAGACGCTGGGCTGATCTTATGATCATTCCTCTTTCCTGCTAA 347

D5 2/8 TCTGCACATTGACCTGCGGAGGTGGGGCAGACCTCCTGCTGTTGGCTCCTCTCTTGAATGCCAA 33/

[illegible]

408 ATGGGATTATAGCAAAACCGTTGGTTTGTTGACCGGCGATGAACACTTTGATGTGGAAACAA 467

Db 398 CTGGACCGCAGACAGCCCTGCTCTTGCCTCCGCGCATGAACCCCATGTGGAGCA 457

QY 468 TCCTTACAGAGAGCGACCTGTCTGTGTGATGACCTGAATACCCATTAATCCCTCC 527

Db 458 CCCGATCACAGCGCAGCGAGGTRAGACCACTTTCAGGCTTTGGCTATGTGAGATCCCTCG 517

QY 528 CATCAAGAAAGAACTGGCTGTGAGACTACGCTAATGCGCCGAATGGCTGA 578

Db 518 TGTGGCCAAAGAGCTGTGTGGAGATGAAGCTCTCGGGGCGCATGGCTGA 568

RESULT 10

BC014409 1240 bp mRNA linear PRI 19-SEP-2001

LOCUS Homo sapiens, hypothetical protein MDS018, clone MGC:19897

DEFINITION IMAGE:4647763, mRNA, complete cds.

ACCESSION BC014409.1 GI:15680132

VERSION BC014409.1

KEYWORDS MGC.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1240)

AUTHORS Strausberg, R.

JOURNAL Direct Submission

Submitted (17-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdel, Yaron Butlerfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stolt, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Zatta.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 29 Row: n Column: 19

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 11141896.

Location/Qualifiers

1. 1240

/organism="Homo sapiens"

/db_xref="locusid:60490"

/db_xref="taxon:9606"

/clone="MGC:19897 IMAGE:4647763"

/tissue_type="Eye, retinoblastoma"

/clone_lib="NIH_MGC_16"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

71. 685

/codon_start=1

/product="hypothetical protein MDS018"

/protein_id="AAH14409.1"

/db_xref="GI:15680133"

/translation="MEPRASCPAAAPLMERKFFHYLVGYTSVAALKLPLVSKLLDIP

GLEAVAYTTERAKHFEVSPDITVLYSDADDEMMKSRSDPLVLDLRRADLLVAP

LDANTLCKVASGICDNLITCYMRAMDSKPLLFPCPAANTAMWEHPITRAOVDOLKARG

YVEIPCAKVLKVCDEGIGAMAEEVGTIVDKVKEVLFQHSFGQGS"

CDS

BASE COUNT 313 a 321 c 356 g 250 t

ORIGIN

Query Match 18.5%; Score 120.6; DB 9; Length 1240;

Best Local Similarity 59.0%; Pred. No. 2.5e-25;

Matches 207; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 228 TGTGACTCTATACAGATGAAGATGAATGCTAGCTGGAACAAGATTGATCCCGT 287

Db 268 TGTGACCCCTCTACAGCAGACGCTGATGAATGGAGATGGGAAGACCCGCTGACCCAGT 327

QY 288 TCTTCATATCAGACTCAGACGCTGGCTGATGCTATGATCATCTGCTCTTCTGCTAA 347

Db 328 TCTGCACATTGTGACCTGCGAGGTGGCAGACCTCTCTGTGGCTCTCTTGTGAAGCCAA 387

QY 348 CACATTAGCAAGATTGCTGTGGTGTGATGATGATATGATGACATGATGATGAAGAC 407

Db 388 CACTCTGGGGAGAGTGGCCAGTGCATCTGTGACAACTTGCTTACTCTGCTATCGGGGC 447

QY 408 ATGGATTATACCAACCGCTGTTGTTGACACCGCGCATGAACACTTGTGATGAGACAA 467

Db 448 CTGGACCGCAGCAGACCCCTGCTCTTGCCTCCGCGCATGAACACCGCATGTGGAGACA 507

QY 468 TCCTTACAGAGAGCGACCTTGTCTGCTGTGATGATGATGATGATGATGATGATGATG 527

Db 508 CCCGATCACAGCGCAGCGAGTGAACCACTCAAGCCCTTGTGCTATGTCGAGATCCCTG 567

QY 528 CATCAAGAAAGAACTGGCTGTGAGACTACGCTAATGCGCCGAATGGCTGA 578

Db 568 TGTGGCCAAAGAGCTGTGTGGAGATGAAGCTCTCGGGGCGCATGGCTGA 618

RESULT 11

AK027491

LOCUS Homo sapiens CDNA FLJ14585 fis, clone NT2RM4001611, weakly similar to S1S2 PROTEIN.

DEFINITION AK027491

ACCESSION AK027491.1 GI:14042205

VERSION AK027491.1

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens teratocarcinoma cell line:NT2 CDNA to mRNA, clone_lib:NT2RM4 clone:NT2RM4001611.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosofiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahata,K., Masuko,Y., Ninomiya,K. and Iwayanagi,T.

TITLE NEBO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2211)

AUTHORS Isogai,T. and Otsuki,T.

JOURNAL Direct Submission

Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan

(E-mail:genom@helix.co.jp, Tel:01-438-52-3975, Fax:01-438-52-3986)

NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

Location/Qualifiers

1. 2211

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="NT2RM4001611"

/cell_line="NT2"

FEATURES

SOURCE

KEYWORDS acetyltransferase complex subunit; actin-interacting protein; arcl family; fat1; glutaredoxin; phosphoethanolamine cytidyltransferase; ribosomal protein l44; thioltransferase; thymidylate synthase-like DNA metabolism protein; vacuolar sorting protein; yeast bud6/Aip3p homolog.

SOURCE Schizosaccharomyces pombe.

ORGANISM Schizosaccharomyces pombe Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.

REFERENCE 1 (bases 1 to 12855) Murphy, L., Harris, D., Barrell, B.G., Rajandream, M.A. and Lyne, M.H. Direct Submission Submitted (09-AUG-1999) Schizosaccharomyces pombe chromosome I sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrell@sanger.ac.uk

COMMENTS Notes: Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/S_pombe)
Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPAC5H10.01c, SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the overlapping sections once, or longer, because we only sequence a small overlap between neighbouring submissions.
Cosmid c15E1 is overlapped at the 5' end by cosmid c15A10.
Location/Qualifiers
1. .12855
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/map="ICEN"
/clone="cosmid c15E1"
1. .1093
/gene="SPAC15E1.01"
/note="SPAC15A10.15; fat1"
1. .1093
/partial
/gene="fat1"
/note="SPAC15E1.01, len:>363"
/codon_start=2
/label=fat1
/product="yeast actin-interacting protein Bud6/Aip3p homolog"
/protein_id="CAB52420.1"
/db_xref="GI:5725406"
/translation="NVEKFEVDVQLEMRSDVLRKVRGDLQDTQSOLNEESAVLEK RIQTLERSLEEVPTNWKQWERELNAIVQDEFLDSHTVLISDLKRDLSALSTVLSN SAIKELAKRSIKSKPLTKATESEIIGRQDIOLEVLINLKPDSEARLQALIESSEL OKRRLQVDEFSKEVTEFVENEKLINKGAEADRIITIDQEKVRKTIAMDPESSKR NGKNGSFISSPTVIDEHNPDNSAKATVAELIDYGSQVTENPMLERSLAPKPL KQPPNITETPIVSTAHEHDEQTPPSKYSNNRVSSSDTYFENTDLKIDNNVMSKY THVHRHDTISTDDYDAEDADVEETSLT"

misc_feature 1. .35
/gene="SPAC15E1.01"
/note="nominal overlap with cosmid SPAC15A10, EM:297208 S. pombe chromosome 1"
complement(1440..1886)
/gene="SPAC15E1.02c"
complement(1440..1886)

gene

CDS

/gene="SPAC15E1.02c"
/note="SPAC15E1.02c, len:148"
/codon_start=1
/label=SPAC15E1.02c
/product="hypothetical protein"
/protein_id="CAB52421.1"
/db_xref="GI:5725407"
/translation="MPSYHMPYVRPSAILGLAFDHAASFVYGPLMGDIWRKAMST APHYVDAGSKTKKVAATYVNSFLSTAVOSYSIALLOLTGTVALKGAFFGLYVFGA SGLPDVDYMFTESSRGTPYILVKTISLSVYKSGVALIGVVR1"
join(2639..2938,3129..3149)
/gene="SPAC15E1.03"
join(2639..2938,3129..3149)
/note="SPAC15E1.03, len:106, SIMILARITY:Phaffia rhodozyma, RL44-PHARH, 60s ribosomal protein l44, (105 aa), fasta scores: opt: 608, E():0, (81.98 identity in 105 aa)"
/codon_start=1
/label=SPAC15E1.03
/product="60s ribosomal protein l44"
/protein_id="CAB52422.1"
/db_xref="GI:5725408"
/translation="MWNIPKTRKTYCGRKNCRKTYVRVYQYKKGPSKLAQGRRD RKQSGFGQTRPVPHKRAKRVKVVRLQECVSKYKNQVLAKRCKHFLGGEKTKGA A10P"
2693..2923
/gene="SPAC15E1.03"
/note="match to PF00935 Ribosomal_L44, Ribosomal protein L44, Score 174.98"
2939..3128
/gene="SPAC15E1.03"
/note="confirmed intron"
2939..2944
/gene="SPAC15E1.03"
/note="gtatgt, splice donor sequence"
3113..3128
/gene="SPAC15E1.03"
/note="ctacaacatcttag, splice branch and acceptor"
3963..5840
/gene="SPAC15E1.04"
3963..5840
/gene="SPAC15E1.04"
/note="SPAC15E1.04, len:625, N-term SIMILARITY:Candida tropicalis, SIS2_CANTR, SIS2 PROTEIN (HALOTOLERANCE PROTEIN HAL3), (531 aa), fasta scores: opt: 613, E(): 1.3e-29, C-term SIMILARITY:Saccharomyces cerevisiae, TYSY_YFASR, thymidylate synthase, (304 aa), fasta scores: opt: 1403, E():0, (65.7% identity in 303 aa)"
/codon_start=1
/label=SPAC15E1.04
/product="thymidylate synthase-like DNA metabolism protein"
/protein_id="CAB52423.1"
/db_xref="GI:5725409"
/translation="MSQPLHARFAFRAVKNPMLKEKRLQTSKYHILVAATGSAVAIL KTLIVLSLTLYKGVQVVLTPDARNFVEKEDLTALGVVYNNADDMKWDGIECP1 THIELRRMHLILAPLSANTRMAKMAKGLCNLTSLIRAWAPLKPIILPAMATYLL TNPITQEHLSAISRIYNSERIMPIETEVACGDIQGMGMEMRIVGRVADKIOLEK SVLPNAVKNIDGDDDSSEQTAAFEEDVDDDDVDVDDNEOSNRIETSANADITPKRS LIPSTTESISKDHETSQAPLGSESVPQTQASENVYTPPEPVPTTSSEVRNTEEOYL NLIRYIIEENGSRPDRGTGTRSVFAPPOLRFSLRNNTLPILTKRVFLKGVLEELIM F1HDDYTNVHLSEKGIHIMDNGSRERLDSGRVGDIGPIYGGOMHFGQAVYD CDPTVTAKGVYLOAVSTILKLNRYDRITLISANPLAIPEMALPCHICQAFVYSP CKRGKQOLSMMYGRADMDLGVPNVIASISLTTHIAHMGYEAEPHYVMGDCHI YNDHLHLQTOLEKRVKAFKFLFKRAKDKIGSIDFSVDDFAVEGYNIPGIKMKHS V"

misc_feature 4944..5837
/gene="SPAC15E1.04"
/note="match to PF00303 thymidylat_synth, Thymidylate synthase, Score 685.78"
complement(join(6052..7082,7138..7204))
/gene="SPAC15E1.05c"
complement(join(6052..7082,7138..7204))

gene

CDS

Db	4375	CTCTTATACGTGACCTGGGCCCCCTTTAAAAACCAATTCCTCTGGCTCCCGCAATGAATFACCT	4434
Qy	4455	TGATGTGGAAACATCCTTTACAGAGACGGCACCTTGCTCTGGTATGAACTTGGAAATCA	514
Db	4435	TGATGTGGACATAATCTTAACTACAGGAACATTTGAGGGCCTATTGACGAATATACAGAA	4494
Qy	515	-----CCCTAATTCCTCCCATCAAGAAAGAACTGGCCCTGTGGAGCTAACGTAATGGCG	568
Db	4495	ATAGTGAATTTATTCATGCACAAATGCAGAAAGTGGCTGTGGTGAATATATTGGCATGGGGC	4554
Qy	569	CAATGGAGTGA	578
Db	4555	GTATGGCCGA	4564
RESULT 14			
SCYKR072C/c			
LOCUS	SCYKR072C	2206 bp	DNA linear
DEFINITION	5. cerevisiae chromosome XI reading frame ORF YKR072c.		
ACCESSION	Z28297.Y13137		
VERSION	Z28297.1	GI:486544	
KEYWORDS	.		
SOURCE	Saccharomyces cerevisiae.		
ORGANISM	Saccharomyces cerevisiae		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
JOURNAL	1 (bases 1 to 2206)		
REFERENCE	Pohl,T.M. and Pohl,F.M.		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 2206)		
JOURNAL	MPS.		
FEATURES	Direct Submission		
source	Submitted (09-MAY-1994) data collected by MIPS on behalf of the European yeast chromosome XI sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@hpmic.mips.biochem.mpg.de		
location/Qualifiers			
1..2206			
/organism="Saccharomyces cerevisiae"			
/db_xref="taxon:4932"			
/chromosome="XI"			
/complement(455..2143)			
/gene="SIS2"			
complement(455..2143)			
/gene="SIS2"			
/note="ORF YKR072c"			
/codon_start=1			
/protein_id="CAA82151.1"			
/db_xref="GI:486545"			
/db_xref="SMISS-PROT:P36024"			
/translation="MTFAVATSGKQDADHNQSTCEPSSRGQKEILLIDHEAKKDKDS			
INSVSGROSISTPLSNATTTTTTSTINAKTSSAAVSNTPPELKRKYATPESDLQK			
QOKQDSLTQLQNDSEKTPSPNSAPVSNIPGHHAVIPNTNTSRITQLSGSPLVNE			
MKDQPKKDSALKIVDTMKDPDKIMATSTPISRNNKVTAKAPSTLRKEDAQDAN			
NVNSQIIVRSTPEETPVKQSVPSIIIPKREMSMLDRLPDQDGLHYLFQATGSLSV			
EKIPMKIKLEIEYGRDRISIOVILLOSATQFFQRTKTKIKSEKLNKMSQESTP			
ATPTPTPGCCNMAOVVELRPHIOLMTDDQEDMDAKRKQRPDVHLIELRRADILVBAR			
LITATLSKIALGLCDNLILTSTRMANSYPIILAPSNVSSFTFNSMTRKLOLTKEEK			
SWATVFKPSEKVVINDITIGGMMDMWEYVKLTVMKLGITPKNNEEDDEDEEDD			
DEEDDTEDKNNNNDDDDDDDDDDDDDDDDDDDDDDDDDEDEDEAETPGIIDKHQ"			
BASE COUNT	545 a	435 c	443 g 783 t
ORIGIN			
Query Match	15.3%	Score 99.8;	DB 8; Length 2206;
Best Local Similarity	60.5%;	Pred. No. 5.5e-19;	
Matches 164; Conservative	0;	Mismatches 107;	Indels 0; Gaps 0;
Qy	218	TACCCAGAAAGTGCACCTCTATACAGATGAAGTAAAGTGTAGCTGGAAGAAAGATTG	277
Db	1086	TACCCCAATATTTAACTATATGACCGACCAAGATGAATGGAGCCATGGAACAGCAA	1027
Qy	278	GTGATCCGCTTCATATCGAGCTCAACAGCGTGGCGTGATGTTATGATCATTTGCTCCTT	337

Db 1026 CTGATCCTGACTTCAATATAGAACTACGTCGCGGCGGATATACGTGTAAGCTCCAT 967
 QY 338 TGTCTGCTAACACATTAAGCCAAAGATTGCTGGCTATGATATCTATTTGACATGTA 397
 Db 966 TAACTGCAACACATTAATCAAAAATGCTTTGGGCTGTGGCAGCATCTTCTGACAACTG 907
 QY 398 TAGTAAGACATGCGGATTAATAGCAACCGTTGTTTGTGCACCGCGCATGAACACTTTGA 457
 Db 906 TCATTAGAGCTTGGAAATCCGAGTATCCCATTTCTTTGGCACCTTCTATGCTGACTAGCA 847
 QY 458 TGTGGAACAATCCTTTTCACAGACGGCACT 488
 Db 846 CTTTAAATCTATGATGACAAAGAAGCACT 816

RESULT 15
 SCU01878

LOCUS SCU01878 4775 bp DNA linear PLN 28-OCT-1995
 DEFINITION Saccharomyces cerevisiae SIS2 gene, complete cds.
 ACCESSION U01878
 VERSION U01878.1 GI:430983
 KEYWORDS
 SOURCE Saccharomyces cerevisiae.
 ORGANISM Saccharomyces cerevisiae.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
 AUTHORS 1 (bases 1 to 4775)
 TITLE Di Como, C.J., Bose, R. and Arndt, K.T.
 JOURNAL Overexpression of SIS2, which contains an extremely acidic region,
 MEDLINE increases the expression of SWI4 and CLN2 in sil4 mutants
 PUBMED Genetics 139 (1), 95-107 (1995)
 7705654

REFERENCE
 AUTHORS 2 (bases 1 to 4775)
 TITLE Di Como, C.J.
 JOURNAL Direct Submission
 TITLE Submitted (16-SEP-1993) Charles J. Di Como, Delbruck, Cold Spring
 Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY
 11724-2212, USA

FEATURES
 source Location/Qualifiers
 1. 4775
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /chromosome="11"
 /map="5cm from MET1"
 291..1214
 /note="product similar to that of open reading frame next
 to CKA2 (chromosome 15)"
 /codon_start=1
 /product="unknown"
 /protein_id="AA079998.1"
 /db_xref="GI:430985"
 /translation="MPSIDPITISYONIMLLDNNTYNNKPAIDYFHHFNDASLEISA
 SWILLKMRKHKLRLPSCSEEDYLDINMTLVRLHCLMRRLSTNHGLONSKNPLS
 INMKETDVTVLGPDLTINIDSNENELSPVQNDKOTKLNKLSALKRNTECWTEK
 DEINASIESNDNALVYLEDISCSBSYDSHTSIFDQSTCTKISSIDEDSEDLNNEK
 EOPPRKLFQNAVVKREIDSKGTRESLIININDIHSRHHRRHHRRHHHHHONSIS
 DETKEAHYFESNTEFCTMEDIFYRNQVVF"
 complement(1365..1793)
 /codon_start=1
 /product="unknown"
 /protein_id="AA079999.1"
 /db_xref="GI:430986"
 /translation="MIMTMEPLTSSNARPLKNSYITLRVLSPEYRWKNTVLHDYDL
 ADKTADLFDNVLNKIONEGSFRFRNPKFTDTLKITTHAGSKTVNLVINFDHDDMT
 LDIENDKKLFEYGIENETELISFNKEDYLFKENPEK"
 2919..4607
 /gene="SIS2"
 2919..4607
 /gene="SIS2"
 /codon_start=1
 /protein_id="AA080000.1"
 /db_xref="GI:430984"

/translation="MRAVASTSGKODADHNOSEICPRFSRQKEIILLDHEDAKKDSI
 INSPVSGROSISPTLSNAATTTTTSKINNAAGTSGAVSNTPDEPLKRPVAVTSDLKQ
 OOKODSLTOLKNDSEPTSPNSNAPVSNISPGHAAVIPHNTNSRTTOLSGSPLVNE
 MKDYDPRKDSALKIVDTMKPKDITMATSTPSRNNKVTAKAPISITLREKEDODAN
 NWSQOINVRSTPEETPYKQSVISITIKRNSKRLDRLQDQCKLVLGAGCSISV
 FKIPMTKLEETIGRDRISIOVLTOSATOFEORRTKIKISSEKLNMSQYESTP
 ATPVTPPGOCNMAQVVELPPIQLMTDQEDMWKORTDPLVHILRRMADILVAP
 LTANTLSKIALGLCDNLITSYIRAMNSYPLILASVSSFTNSMTKROLQITKEEM
 SWTVFPRSEKVDINDGIDIGAGMDNNEIVNKIVMKLGGYPRKNEEEDDEDEEDD
 DEEDTIDKNNNNDDDDDDDDDDDDDDDDDDDDDDDDDEDEDEAEFTGIIIDKHQ"

BASE COUNT 1561 a 977 c 916 g 1321 t
 ORIGIN

Query Match 15.3%; Score 99.8; DB 8; Length 4775;
 Best Local Similarity 60.5%; Pred. No. 6e-19;
 Matches 164; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 218 TACTCTGACATGTGACTCTCTATACAGATGAATGCTTACGCTGGAACAAGATTG 277
 Db 3976 TACCCACACATATTCACATATGACCGACCAAGATGAATGGAGCATGAACAACGAA 4035
 QY 278 GTGATCCGCTTCTCATATGACAGCTCAGACGCTGGCGTGAATGATATCAFTGCTCCT 337
 Db 4036 CTGATCCTGATCTCATATAGAACTACGTCGCGGCGGATATACTGTAAGCTCCAT 4095
 QY 338 TGTCTGCTAACACATTAAGCCAAAGATTGCTGGCTATGATATCTATTTGACATGTA 397
 Db 4096 TAACTGCAACACATTAATCAAAAATGCTTTGGGCTGTGGCAGCATCTTCTGACAACTG 4155
 QY 398 TAGTAAGACATGCGGATTAATAGCAACCGTTGTTTGTGCACCGCGCATGAACACTTTGA 457
 Db 4156 TCATTAGAGCTTGGAAATCCGAGTATCCCATTTCTTTGGCACCTTCTATGCTGACTAGCA 4215
 QY 458 TGTGGAACAATCCTTTTCACAGACGGCACT 488
 Db 4216 CTTTAAATCTATGATGACAAAGAAGCACT 4246

Search completed: December 1, 2002, 20:15:38
 Job time : 3393 secs

Db 63 TWWKTRPDPVLIHIELRRMADVFIAPLPLANTLAKIANGICDNLITSVIRAWDTNKPILFC 122

Oy 147 PAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSLIYTVR 202

Db 123 PAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSLIYTVR 178

RESULT 2

Oy 09AVS8 PRELIMINARY: PRT: 120 AA.

AC 09AVS8:

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative halotolerance protein HAL3 homolog (Fragment).

GN PPA0038.

OS Picea abies (Norway spruce) (Picea excelsa).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

OX NCBI_TaxID=3329;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-KARST;

RA Schubert R., Mueller-Starck G., Riegel R.;

RT "Development of EST-PCR markers and monitoring their intrapopulation

RL genetic variation in Picea abies (L.) Karst.;"

DR EMBL: AJ271130; CAC27336.1; -.

DR HSSP: Q9SWE5; 1E20

DR InterPro: IPR003382; Flavoprotein.

DR Pfam: PF02441; Flavoprotein; 1.

FT NON_TER 1

SO SEQUENCE 120 AA; 13064 MW; 416423419A49FDA CRC64;

Query Match 41.5%; Score 475; DB 10; Length 120;

Best Local Similarity 79.4%; Pred. No. 5, 1e-40;

Matches 85; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Oy 96 VLIHIELRRMADVIAPLPLANTLAKIAGLCLDNLITCIVRAMDYSKPLFVAPAMNTLMNN 155

Db 6 VLIHIELRRMADVIAPLPLANTLAKIAGLCLDNLITCIVRAMDYSKPLFVAPAMNTLMNN 65

Oy 156 NPTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSLIYTVR 202

Db 66 NPTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSLIYTVR 112

RESULT 3

Oy 096SX0 PRELIMINARY: PRT: 204 AA.

AC 096SX0:

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE CDNA FLJ1585 fis, clone NT2RM4001611, weakly similar to SIS2

DE protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Isogai T., Oota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,

RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,

RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,

RA Ninomiya K., Iwayanagi T.;

RT "NDO human cDNA sequencing project";

RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AK027491; BAB5151.1; -.

DR InterPro: IPR003382; Flavoprotein.

DR Pfam: PF02441; Flavoprotein; 1.

SO SEQUENCE 204 AA; 22429 MW; 98808FA0207C6B66 CRC64;

Query Match 33.9%; Score 445; DB 4; Length 204;

Best Local Similarity 48.4%; Pred. No. 1e-36;

Matches 90; Conservative 26; Mismatches 64; Indels 6; Gaps 3;

Oy 26 RKPRILLASGSVASIKFSNLCHEFSE--WAEYKAVASKSSLNFDKPSLPQNTVLTDE 83

Db 16 RKPHVLGVGTGSVAALKLPVLVSKLDIPGLEVSVYTERAKHFSPQDIP--VTLYSDA 73

Oy 84 DEWSSWNKIGDPPVLIHIELRRMADVIAPLPLANTLAKIAGLCLDNLITCIVRAMDYSKPL 143

Db 74 DEWEMKMSRSDPYLHIDLRRMADLIVAPLDANTLGKVASGICDNLITCIVRAMDYSKPL 133

Oy 144 EVAAPAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSLIYTVR- 202

Db 134 LFCPAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEVGTIVDKVE 193

Oy 203 -LEWES 207

Db 194 VLFQHS 199

RESULT 4

Oy 096CD2 PRELIMINARY: PRT: 204 AA.

AC 096CD2:

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Hypothetical 22.4 kDa protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-EYE;

RA Strausberg R.;

RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: BC014409; AAH14409.1; -.

DR InterPro: IPR003382; Flavoprotein.

DR Pfam: PF02441; Flavoprotein; 1.

KW Hypothetical protein.

SO SEQUENCE 204 AA; 22413 MW; 9899683486C32A54 CRC64;

Query Match 38.8%; Score 444; DB 4; Length 204;

Best Local Similarity 48.4%; Pred. No. 1, 3e-36;

Matches 90; Conservative 26; Mismatches 64; Indels 6; Gaps 3;

Oy 26 RKPRILLASGSVASIKFSNLCHEFSE--WAEYKAVASKSSLNFDKPSLPQNTVLTDE 83

Db 16 RKPHVLGVGTGSVAALKLPVLVSKLDIPGLEVSVYTERAKHFSPQDIP--VTLYSDA 73

Oy 84 DEWSSWNKIGDPPVLIHIELRRMADVIAPLPLANTLAKIAGLCLDNLITCIVRAMDYSKPL 143

Db 74 DEWEMKMSRSDPYLHIDLRRMADLIVAPLDANTLGKVASGICDNLITCIVRAMDYSKPL 133

Oy 144 EVAAPAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEPSLIYTVR- 202

Db 134 LFCPAMNTLMNNPFTERRHLVLLDELGITLIPPIKKKLACGDBGNGAMAEVGTIVDKVE 193

Oy 203 -LEWES 207

Db 194 VLFQHS 199

RESULT 5

Oy 09UT17 PRELIMINARY: PRT: 625 AA.

AC 09UT17:

DT 01-MAY-2000 (TREMBLrel. 13, Created)


```
RESULT 8
09HC17 PRELIMINARY: PRT: 127 AA.
AC 09HC17.
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE MDS018.
GN MDS018.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.:
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT Myelodysplastic Syndromes patient."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF182419; AAG14955.1; -.
DR HSSP: Q9SWE5; 1E20.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
SQ SEQUENCE 127 AA; 14069 MW; 562D659DE77C0AF5 CRC64;

Query Match
Best Local Similarity 56.2%; Score 361; DB 4; Length 127;
Matches 68; Conservative 14; Mismatches 37; Indels 2; Gaps 1;

QY 89 WNKIGDPVLHIELRRADVMIIAPLSANTLAKIAGLCDNLTCIVRAMDYSKPLFVAPA 148
DB 2 WKSRSPPVLHIDLRMADLLVAPLADNTLGKVASGICNLLTCVARMDSKPLFPCPA 61
QY 149 MNTLMNNPFTEHVLVLDLDELITLIPPIKKIACGDYNGMAEFLSYSTR--LFWE 206
DB 62 MNTAMEHEITAOQVQLKAFGVLEIPCAKRLVCGDEGMAAEVGTIVDKVEVLFQH 121
QY 207 S 207
DB 122 S 122

RESULT 9
008438 PRELIMINARY: PRT: 674 AA.
AC 008438: 000019;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ORF_YOR054C.
GN YOR054C OR YOR529-05.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE OF 673-674 FROM N.A.
RA Landt O., Hiesel R., Unselid M.:
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bohn C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,
RA Valens M.:
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS:
RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=97279235; PubMed=9133743;
RA Valens M., Bohn C., Daignan-Fornier B., Dang V., Bolotin-Fukuhara M.;
```

```
RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals the
RT presence of two tRNAs and 24 new open reading frames."
RL Yeast 13:379-390(1997).
DR EMBL: 274961; CAA99246.1; -.
DR EMBL: 270678; CAA94539.1; -.
DR HSSP: Q9SWE5; 1E20.
DR SGD: S0005580; YOR054C.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
SQ SEQUENCE 674 AA; 73648 MW; A97623473C6BF605 CRC64;

Query Match
Best Local Similarity 36.6%; Score 323; DB 3; Length 674;
Matches 71; Conservative 40; Mismatches 57; Indels 26; Gaps 5;

QY 11 NKEIKMNEVDPTKRPRIILAASGSVASIKFSNCHCFSEMAEYKAAVASKSSLNFVOK 70
DB 395 NKQEEKQMA--STGEPSTL-----GGSRTYSNNSNVSOHPQLE----- 433
QY 71 PSLPQWVLTYTEDEWSSNNKIGDPVLHIELRRADVMIIAPLSANTLAKIAGLCDNLL 130
DB 434 --LPAHIOFWTDDQDQVRRQRTDPLHIELRRADILVAPLTANTLAKIALGCDNLL 491
QY 131 TCIVRAMDYSKPLFVAPAMNTLMNNPFTEHVLVLDL--LGITLIPPIKKIAC--GDYG 187
DB 492 TSVIRAMNPFPIFLAPSMGSGTFNSIMTKHFRITIOEEMPVWTVKPSKVGINGDIG 551
QY 188 NGMAEPLSYSTRV 201
DB 552 LSGMDANETVGTKI 565

RESULT 10
066997 PRELIMINARY: PRT: 388 AA.
AC 066997;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pantothenate metabolism flavoprotein.
GN DFP OR AQ_815.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.:
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
RL EMBL: AE000708; AAC06944.1; -.
DR HSSP: Q9SWE5; 1E20.
DR InterPro: IPR005252; Cons.hypoth521.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
DR TIGRPFMS: TIGR00521; dfp; 1.
KW Complete proteome.
SQ SEQUENCE 388 AA; 43539 MW; 682BA44B64524D0B CRC64;

Query Match
Best Local Similarity 24.5%; Score 280.5; DB 16; Length 388;
Matches 70; Conservative 26; Mismatches 55; Indels 19; Gaps 7;

QY 30 ILLAASGSVASIKFSNCHCFSEMA----EVKAVASKSSLNFVOKPSLPQNT---LYTD 82
DB 3 ILIGVCGGIASRY---VCELVRELKRGHSVKITLIPPAKFM-SPLTROTISGNKAYTD 58
QY 83 EDWSSNNKIGDPVLHIELRRADVMIIAPLSANTLAKIAGLCDNLLTCIVRAMDYSKP 142
```

```

DB 59 KD-----WEE--EPLAHINARNADVFLIAPATANTIAKLANGISDNLTTTILA--YGRP 110
OY 143 LEVAPAMNTLMNNPTEHRLVLLDELGITLPIPKKILACGDYNGAMA 192
DB 111 LVAAPAMNTVMYKSPSTQENLKKEMGHVILIEPEGVLACKEVEGSKLA 160

RESULT 11
OY 099210 PRELIMINARY; PRT: 181 AA.
AC 099210:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE putative DNA/pantothenate metabolism flavoprotein.
GN DPF OR SPY1222.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RA MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., Keshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Pirmeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RA *Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006562; AK34081.1; -.
DR HSSP; Q9SWE5; I220.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR Complete proteome.
SO SEQUENCE 181 AA; 19443 MW; BABA9B727899D7F CRC64;

Query Match 24.0%; Score 274.5; DB 16; Length 181;
Best Local Similarity 35.5%; Pred. No. 1.3e-19;
Matches 65; Conservative 37; Mismatches 68; Indels 13; Gaps 4;

OY 30 ILLAASGSVASIKFSNLCHESEMA-EVKAVASKSLNFVDRPS---LPQNTLYTDEE 85
DB 5 ILLAASGSVASIKAKADLTQSRLKIGDVHIIIMTQATQFTLPLQVLSIKMIHLDVME 64
OY 86 WSSWNKIGDP--VLHIELRRMADVMIIAPLSANTLAKIAGLCNDLTCITVAMVDSKPL 143
DB 65 -----HDPKVINIHELAKRTDLEIVAPASANTIAHLAYGADNLVTSVALALPATTPK 117
OY 144 EVAPAMNTLMNNPTEHRLVLLDELGITLPIPKKILACGDYNGAMAEPSLIVSTVRL 203
DB 118 LVAAPAMNTVMYKSPSTQENLKKEMGHVILIEPEGVLACKEVEGSKLA 177
OY 204 FWE 206
DB 178 IWK 180

RESULT 12
OY 08YSH4 PRELIMINARY; PRT: 410 AA.
AC 08YSH4:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pantothenate metabolism flavoprotein.
GN ALI3111.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN 11
RP SEQUENCE FROM N.A.
SO MEDLINE=21595285; PubMed=11759840;

```

```

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT *Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003591; BAB74810.1; -.
DR InterPro; IPR005252; Cons.hypoth521.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; ddp; 1.
DR Complete proteome.
SO SEQUENCE 410 AA; 44563 MW; 1593605E9608F126 CRC64;

Query Match 23.6%; Score 270.5; DB 16; Length 410;
Best Local Similarity 38.0%; Pred. No. 8.9e-19;
Matches 76; Conservative 29; Mismatches 74; Indels 21; Gaps 7;

```

```

OY 26 RKPRILLAASGSVASIK----FSNLCHESEMAEVAKAVASKSLNFVDRPSLPQNTL-- 79
DB 11 KKSRLVAVANGGIIATKVEEVSNL---FKSGEVAVILTRSOEFTITPLTIS---TLSR 64
OY 80 ----YTDEDEWSSWNKIGDPEVLHIELRRMADVMIIAPLSANTLAKIAGLCNDLTCITVRA 136
DB 65 HNAVYDDDEFWQSH---SRPLHIELGEMADLLVIAPLANTLAKIAGLGMADNLTFVTVA 121
OY 137 WDSKRLFAAPAMNTLMNNPTEHRLVLLDELGITLPIPKKILACGDYNGAMAEPS 195
DB 122 --SNCPALLAPAMNTDMWQLTYVRMWOQLTDSRYHGIGTAGSLACDVGAGRMABEP 179
OY 196 LIYSIVRLFWESQARKRDG 215
DB 180 ELIATYQSLHTQGRDLAG 199

RESULT 13
OY 097010 PRELIMINARY; PRT: 183 AA.
AC 097010:
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Flavoprotein.
GN SPI231.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT *Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007423; AAK75337.1; -.
DR TIGR; SPI231; -.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR Complete proteome.
SO SEQUENCE 183 AA; 19866 MW; 16DBB4EFB1CA18D8 CRC64;

Query Match 23.1%; Score 264.5; DB 16; Length 183;

```

Best Local Similarity 36.0%; Pred. No. 1.3e-18;
Matches 64; Conservative 39; Mismatches 64; Indels 11; Gaps 4;

OY 30 ILLAAGSVASIKFSNLCHCFSEMA-EVKAVASKSSLNPFVDRS---LPON-VTLTDED 84
DB 4 ILLAAGSVASIKFSNLCHCFSEMA-EVKAVASKSSLNPFVDRS---LPON-VTLTDED 63
OY 85 EWSNNKIDDPVLIHLELRWADVMIIAPLSANTLAKIAGLCIDNLTCTIVRAMDYSKPLF 144
DB 64 PVP-----DQVNHIELGKKADLFIVVPATANTIAKLHGFADNMVTSIALALPSHPKL 117
OY 145 VARAANTLMNNPETERHVLDELGITLPIPIKKKLACGDYNGAMAEPSLIYSTRV 202
DB 118 IAPAMNTKMYDHPVTONNKTETTYGYOLIAPEKESLACGDHGRGALADLTILLERIK 175

RESULT 14

OY 09X3X4 PRELIMINARY; PRT: 417 AA.

AC 09X3X4;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE PanOthenate metabolism flavoprotein.
GN DFP.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RT Lee H.S., Kang H.S.;
RL "Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis ZM4.";
EMBL: AF088896; AAD21550.1; -;
DR HSSP; Q9SWES; 1E20.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMS; TIGR00521; dfp; 1.
SQ SEQUENCE 417 AA; 45382 MW; 492P6585E5379138 CRC64;

Query Match 22.8%; Score 261; DB 2; Length 417;
Best Local Similarity 35.2%; Pred. No. 8.2e-18;
Matches 64; Conservative 39; Mismatches 73; Indels 6; Gaps 4;

OY 29 RILLAAGSVASIKFSNLCHCF-SEMAEYKAVASKSSLNPFVDRS---LPQNVTLTYTDEDEMS 87
DB 9 RILLVSGSIAAIKAPDITIRLFRRKKADIRCLITTKGANFTPLALASLSGNFVADMD 68
OY 88 SNKKICDPVLIHLELRWADVMIIAPLSANTLAKIAGLCIDNLTCTIVRAMDYSKPLFVAP 147
DB 69 ESEAS--IRHRLAREADMIIVAPASADFISIKMAHGLANDLASTVLAAD--SPILVAP 124
OY 148 AMNTLMNNPETERHVLDELGITLPIPIKKKLACGDYNGAMAEPSLIYSTRV 206
DB 125 ANNHRRMHHHSATQRNIHQKSDCISIVDPBAGAMACGETIGRLAPEDILLASBSLPAE 184
OY 207 SQ 208
DB 185 EQ 186

RESULT 15

OY 09A254 PRELIMINARY; PRT: 412 AA.

AC 09A254;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE DNA/panOthenate metabolism flavoprotein.
GN CC3712.

OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utecherback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE006029; AAK25674.1; -;
DR HSSP; Q9SWES; 1E20.
DR TIGR; CC3712; -;
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMS; TIGR00521; dfp; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 43105 MW; 9D4D24ED3A218515 CRC64;

Query Match 22.4%; Score 256.5; DB 16; Length 412;
Best Local Similarity 32.1%; Pred. No. 2.3e-17;
Matches 63; Conservative 39; Mismatches 73; Indels 21; Gaps 4;

OY 29 RILLAAGSVASIKFSNLCHCFSE-MAEYKAVASKSSLNPFVDRS-----QNVTL 79
DB 19 RILLVGGGAIAKALLRLRLKAGVAVRPILTKGAAFPVPLSLAALAEKVEDLES 78
OY 80 YNDEDESSNNKIDPVLHLELRWADVMIIAPLSANTLAKIAGLCIDNLTCTIVRAMDY 139
DB 79 LTDEHEMG-----HIELSRSDLVVAVAPATADLIKAAGLACGLASTLLATD- 127
OY 140 SKPLFVAPAMNTLMNNPETERHVLDELGITLPIPIKKKLACGDYNGAMAEPSLIYS 199
DB 128 -KPVLMAPAMNVMWLPATQRNIATLEADGVRFVPEGAMACGEGGGRILAEPALFA 186
OY 200 TVRLFWEQARKOROG 215
DB 187 ATMALEGPAPPLES 202

Search completed: November 25, 2002, 10:15:33
Job time : 35 secs

Run on: November 25, 2002, 10:10:45 ; Search time 12 seconds
OM protein - protein search, using SW model
(without alignments)
750.030 Million cell

```
Title: 1145
Perfect score: 1 EFGTSSFLGCKNKKKKMMME.....
Sequence: 1
```

```

112892 segs, ...
parameters:
11289

```

Minimum DB seq length:	2000000000
Maximum DB seq	

Maximum Match	0%
Minimum Match	100%
Maximum Match	45 summaries
Minimum Match	first 45 summaries

Database: **SwissProt_40**

SPSSPROF 4.0.0

case 1

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

SUMMARIES

Result NO.	Score	Query Match	Length	DB	ID	Description
1	1060	92.6	201	1	HL3B-ARATH	P94063 arabidopsis
2	946	82.6	209	1	HL3B-ARATH	P94063 arabidopsis
3	347	30.3	571	1	YK18-YEAST	Q96956 saccharomyc
4	331.5	29.0	562	1	S1S2-YEAST	Q96956 saccharomyc
5	329.5	28.3	531	1	S1S2-CANTR	P36000 candida tlo
6	323.5	20.9	402	1	DEP-SYNY3	P73881 synechocyst
7	236.5	20.7	400	1	DEP-HAELIN	Q54433 haemophilus
8	217.5	19.0	405	1	DEP-STRMU	P44933 streptococc
9	205.5	17.9	456	1	DEP-ECOLI	Q54433 escherichia
10	184	16.1	390	1	EPID-STAPB	P24288 staphylococ
11	176.5	15.4	403	1	DEP-BOBHU	P30197 borrelia bu
12	135.5	11.8	418	1	DEP-MYCTU	Q51752 mycobacteri
13	109	9.5	206	1	DEP-METUA	P71561 mycobacteri
14	108	9.1	182	1	DEP-METUA	Q58323 methanococ
15	104	9.0	189	1	DEP-METUA	Q57566 methanococ
16	102.5	7.9	755	1	PAAD-PSFEE	P72743 synechocyst
17	90.5	7.8	181	1	PAAD-SYNY3	Q29054 archaeoglob
18	89.5	7.8	423	1	PAAD-METUA	Q93084 pseudomonas
19	89.5	7.8	423	1	PAAD-SYNY3	G93084 pseudomonas
20	89.5	7.8	423	1	PAAD-SYNY3	P49250 klebsiella
21	89	7.8	423	1	PAAD-SYNY3	Q58742 pyrococcus
22	88	7.7	423	1	PAAD-SYNY3	Q58742 pyrococcus
23	87.5	7.6	428	1	PAAD-SYNY3	P38690 saccharomyc
24	85	7.4	428	1	PAAD-SYNY3	Q94404 bacillus su
25	84	7.3	329	1	PAAD-SYNY3	Q58140 methanococ
26	83.5	7.3	192	1	PAAD-SYNY3	Q99143 staphylococ
27	83	7.2	329	1	PAAD-SYNY3	Q99143 staphylococ
28	82	7.2	329	1	PAAD-SYNY3	Q99143 staphylococ
29	81	7.1	187	1	PAAD-SYNY3	Q99143 staphylococ
30	80.5	7.0	187	1	PAAD-SYNY3	Q99143 staphylococ
31	80	7.0	211	1	PAAD-SYNY3	Q99143 staphylococ
32	80	7.0	211	1	PAAD-SYNY3	Q99143 staphylococ
33	80	7.0	211	1	PAAD-SYNY3	Q99143 staphylococ

[illegible]

ALIGNMENTS

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U80192; AAF53106.1; -
 DR EMBL: AC020889; AAF79709.1; ALT_INT.
 DR HSP: Q9SWES; 1E20.
 DR InterPro: IPR003382; Flavoprotein.
 DR Pfam: PF02441; Flavoprotein; 1.
 KW Flavoprotein; FMN.
 SQ SEQUENCE 201 AA; 22415 MW; 7962655E408FA64C CRC64;

Query Match 92.6%; Score 1060; DB 1; Length 201;
 Best Local Similarity 100.0%; Pred. No. 2.7e-96;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEYKAVAKSSLNFDKPSLPON 76
 DB 1 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEYKAVAKSSLNFDKPSLPON 60
 QY 77 VLTLYTDEDESSWNKIGDPVLIETLRMDVMTIAPLSANTLAKINGGLCDNLTCIVRA 136
 DB 61 VLTLYTDEDESSWNKIGDPVLIETLRMDVMTIAPLSANTLAKINGGLCDNLTCIVRA 120
 QY 137 WDYSKPLFVAPAMNTLMNNPFTERRHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 196
 DB 121 WDYSKPLFVAPAMNTLMNNPFTERRHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 180
 QY 197 IYTVRLFWEESQARKORDGTS 217
 DB 181 IYTVRLFWEESQARKORDGTS 201

RESULT 2

HL3A_ARATH STANDARD; PRT; 209 AA.

AC Q9SWES; 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Halotolerance protein Hal3A.
 GN HAL3A OR ARG318030 OR MBG14.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A. TISSUE SPECIFICITY, AND INDUCTION.
 RC STRAIN=cv. Columbia;
 RA MEDLINE=20117504; PubMed=10652125;
 RA Espinosa-Ruiz A., Belles J.M., Serrano R., Cullanez-Macia F.A.;
 RT "Arabidopsis thaliana AtHAL3: a flavoprotein related to salt and
 RT osmotic tolerance and plant growth.";
 RT Plant J. 20:529-539(1999).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=20363099; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RT DNA Res. 7:217-221(2000).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20444410; PubMed=10984663;
 RA Albert A., Martinez-Ripoll M., Espinosa-Ruiz A., Yennush L.,
 RA Cullanez-Macia F.A., Serrano R.;

RT "The X-ray structure of the FMN-binding protein AtHAL3 provides the
 RT structural basis for the activity of a regulatory subunit involved in
 RT signal transduction.";
 RL Structure 8:961-969(2000).
 CC -1- FUNCTION: INVOLVED IN PLANT GROWTH AND SALT AND OSMOTIC TOLERANCE.
 CC MAY CATALYZE THE ALPHA,BETA-DEHYDROGENATION OF A PEPTIDYL CYSTEINE
 CC OF A CELLULAR PARTNER.
 CC -1- COFACTOR: FMN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, SHOOTS, LEAVES, FLOWERS,
 CC DEVELOPING STILLOES AND SEEDS WITH HIGHEST EXPRESSION IN SEED
 CC EMBRYOS AND PHLOEM.
 CC -1- INDUCTION: BY SALT STRESS.
 CC -1- SIMILARITY: SOME, TO YEAST AND C.TROPICALIS SIS2/HAL3 AND YEAST
 CC YKL088W.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF166262; AAD51616.1; -
 DR EMBL: AB026641; BAB01331.1; -
 DR PDB: 1E20; 1I-SEP-00.
 DR InterPro: IPR003382; Flavoprotein.
 DR Pfam: PF02441; Flavoprotein; 1.
 KW Flavoprotein; FMN; 3D-structure.
 SQ SEQUENCE 209 AA; 23355 MW; 3AB1B8364F8E40DE CRC64;

Query Match 82.6%; Score 946; DB 1; Length 209;
 Best Local Similarity 85.6%; Pred. No. 3.7e-85;
 Matches 173; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 16 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEYKAVAKSSLNFDKPSLPQ 75
 DB 8 MNNEVDVTYRKPRILLAAAGSVASIKFSNLCHEFSEMAEYKAVAKSSLNFDKPSLPQ 67
 QY 76 NVTLYTDEDESSWNKIGDPVLIETLRMDVMTIAPLSANTLAKINGGLCDNLTCIVR 135
 DB 68 NVTLYTDEDESSWNKIGDPVLIETLRMDVMTIAPLSANTLAKINGGLCDNLTCIVR 127
 QY 136 AMDYKPLFVAPAMNTLMNNPFTERRHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 195
 DB 128 AMDYKPLFVAPAMNTLMNNPFTERRHLVLDLGTLLIPPIKKKLACG DYGNAMAEP SL 187
 QY 196 LIYTVRLFWEESQARKORDGTS 217
 DB 188 LIYTVRLFWEESQARKORDGTS 209

RESULT 3

YK18_YEAST STANDARD; PRT; 571 AA.

AC P36076; 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 65.2 kDa protein in MIF2-CYT2 intergenic region.
 GN YKL088W.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pohl T.M., Pohl F.M.;
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBI databases.
 CC -1- SIMILARITY: TO YEAST AND C.TROPICALIS SIS2/HAL3. SOME, TO
 CC A.THALIANA HAL3A AND HAL3B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; 228088; CAA81926.1; -
 DR PIR; S37913; S37913.
 DR HSSP; Q9SWE5; 1E20.
 DR SGD; S0001571; YKL088W.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 DR Hypothetical protein.
 DR KW Hypothetical protein.
 FT DOMAIN 508 570 ASP/GLU-RICH (HIGHLY ACIDIC).
 SQ SEQUENCE 571 AA: 65238 MW: 96674C2394EFC6AB CRC64;

Query Match 30.3%; Score 347; DB 1; Length 571;
 Best Local Similarity 37.6%; Pred. No. 3.8e-26;
 Matches 89; Conservative 39; Mismatches 77; Indels 32; Gaps 8;

OY 8 LGCNRIKKMKNMEVDVTRKPR-----LLLAAGSVASIK---FSNIC 47
 DB 274 LSANNSTNR-NKEVITAPTPGPRVPTFEQKEDDKKFHILIGATGSGVATIKVPLIIDKLF 332
 OY 48 HCFSS-EMAEVKAVASKSSLNFEVKPSLPONTLYTDEDEM--SSNKKIG----DPVLEHT 99
 DB 333 KLYGEKISIQILVTRKPAHEFLKGLKMSHVKIMREDMVDAVKNKNTSLNLILHH 392
 OY 100 ELRKADWMIAPLSANTLAKIAGLCNDLITCIVRAMDYSKPLEVAPAMNTLMNNPFT 159
 DB 393 ELRKADWIFLPLASANTLAKLANGLICNNLLTSVYRDMSPLEVPVLIAPAMNTFMYINPWT 452
 OY 160 ERLVLL--DELGITLIPPIKKKLAGDVGNGAMAEPSLIYSTVRLFMFSQARKQD 214
 DB 453 KKHLSLVQDYPFIQVLKPEKVLICGDI GGMGRMTDIVEIVRR-RINEIRKARD 508

RESULT 4
 SIS2_YEAST
 ID SIS2_YEAST STANDARD: PRT: 562 AA.
 AC P36024.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SIS2 protein (Halotolerance protein HAL3).
 GN SIS2 OR HAL3 OR YKR072C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pohl T.M., Pohl F.M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=96009574; PubMed=7565698;
 RA Ferrando A., Kiron S.J., Rios G., Fink G.R., Serrano R.;
 RT "Regulation of cation transport in Saccharomyces cerevisiae by the
 RL salt tolerance gene HAL3.";
 RX Mol. Cell. Biol. 15:5470-5481(1995).
 RN [4]
 RP IDENTIFICATION AS INHIBITORY SUBUNIT OF PP21.
 RX MEDLINE=96301578; PubMed=9636153;

RA de Nadal E., Clotet J., Posas F., Serrano R., Gomez N., Arino J.;
 RT "The yeast halotolerance determinant Halp3 is an inhibitory subunit of
 RT the Ppz3 ser/Thr protein phosphatase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:7357-7362(1998).
 CC -I- FUNCTION: MAY STIMULATE EXPRESSION OF CERTAIN GENES THAT ARE
 CC PERIODICALLY EXPRESSED DURING LATE G1. ALSO MODULATES THE
 CC EXPRESSION OF THE ENA1 ATPASE. INTERACTS WITH THE C-TERMINAL
 CC DOMAIN OF THE SERINE-THREONINE PROTEIN PHOSPHATASE PP21 AND ACTS
 CC AS AN INHIBITORY SUBUNIT OF PP21.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR OR CYTOPLASMIC.
 CC -I- SIMILARITY: TO C.TROPICALIS SIS2/HAL3 AND TO YEAST YKL088W. SOME,
 CC TO A.THALIANA HAL3A AND HAL3B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U01878; AAA80000.1; -
 DR PIR; S38149; S38149.
 DR HSSP; Q9SWE5; 1E20.
 DR SGD; S0001780; SIS2.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 DR KW Protein phosphatase inhibitor; Nuclear protein.
 FT DOMAIN 496 553 ASP/GLU-RICH (HIGHLY ACIDIC).
 SQ SEQUENCE 562 AA: 62478 MW: 19A9A475145DA7AB CRC64;

Query Match 29.0%; Score 331.5; DB 1; Length 562;
 Best Local Similarity 35.6%; Pred. No. 1.2e-24;
 Matches 77; Conservative 41; Mismatches 67; Indels 31; Gaps 6;

OY 2 FETSSFLGCKRIKKMMEVDVTRKPRILLAAAGSVASIKSNCCHCFSEMAEYKAVAS 61
 DB 268 FGATGSLAVFKI-KPMIKLEIYGRDI-----SIQVILQSATQFREQRYTKIKS 320
 OY 62 KSSLN---FVDRPS-----LPQNVITYTDEDESSNNKIGDPVLEHT 100
 DB 321 SEKLKMSQYESTPATPVPTPGQCMAQVLELPPIQLMTDODEDAKORPDPVLEHT 380
 OY 101 LRRMADWMIAPLSANTLAKIAGLCNDLITCIVRAMDYSKPLEVAPAMNTLMNNPFT 160
 DB 381 LRRMADWIVVAPLANTLSKIALGLCNDLITSVIRAMNSYPIILLAPSVSSTPNSMTRK 440
 OY 161 RHLVLLDE--LGITLIPPIKKRIAC-GDYNGAMAE 193
 DB 441 KQLQTIKEMSWVTYFKPKSEKVDINGDIGLGGMMD 476

RESULT 5
 SIS2_CANTR
 ID SIS2_CANTR STANDARD: PRT: 531 AA.
 AC Q12600.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE SIS2 protein (Halotolerance protein HAL3).
 GN SIS2 OR HAL3.
 OS Candida tropicalis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5482;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=NCYC 2512;
 RC MEDLINE=97082501; PubMed=8923737;
 RA Rodriguez P.L., Ali R., Serrano R.;
 RT "Ctdc5p and Cth13p: two putative regulatory proteins from Candida
 RT tropicalis with long acidic domains.";

```

RL Yeast 12:1321-1329(1996).
CC -1- FUNCTION: MAY STIMULATE EXPRESSION OF CERTAIN GENES THAT ARE
CC PERIODICALLY EXPRESSED DURING LATE G1. ALSO MODULATES THE
CC EXPRESSION OF THE ENAI ATPASE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR OR CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: TO YEAST SIS2/HAL3 AND TO YEAST YKL088W. SOME, TO
CC A THALIANA HAL3A AND HAL3B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X88900; CA61362.1; -.
CC DR HSSP: Q9SWE5; IE20.
CC DR InterPro: IPR003382; Flavoprotein.
CC DR Pfam: PF02441; Flavoprotein; 1.
CC FT DOMAIN 233 240 POLY-THR.
CC FT DOMAIN 246 249 POLY-THR.
CC FT DOMAIN 484 513 ASP/GLU-RICH (HIGHLY ACIDIC).
CC SO SEQUENCE 531 AA: 58171 MW; D5FF196BB957B14 CRC64;
CC -----
Query Match 28.3%; Score 324.5; DB 1; Length 531;
Best Local Similarity 36.6%; Pred. No. 5,5e-24;
Matches 78; Conservative 46; Mismatches 60; Indels 29; Gaps 9;
OY 4 TSSPLGCKNTEKKMNEVTYTRKPR-----LLAASG--SVASIKF--SNLCICF-SE 52
DB 240 TNSNSNTTTCGEOSNSND--PRLPODDGKPHVLIGVCGALSGVKYKLIIVNKLLEYTSD 297
OY 53 WAEVKAASKSLNFWKDSLPDPO-----NVTLYTDEDESSNKKIGDPYLHIELRR 103
DB 298 KTSIOVILTKSSENL-----LPETLVNLENVKKYKRWITIDERTTKTLDPLHIELRR 353
OY 104 WADVMIAPLSANTLAKIAGLGDNLITCIVRAMDYSKPLFVAPAMNTLMNPNTERHL 163
DB 354 WADILLVCLPTANTLAKISLGICDNLNVIRAMNSYPTILLAPAMDHSYSSTTKROL 413
OY 164 VLL--DELGITLIPPIKKRL-ACGDYGCNMAAE 193
DB 414 RLADDMPIEVLKPLEKVFSGDYGMD 446
CC -----
RESULT 6
DFF_SYNY3 STANDARD: PRT; 402 AA.
AC P73881;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA/pantothenate metabolism Flavoprotein homolog.
GN DFP OR SL0250.
OS Synecchocystis sp. (strain PCC 6803).
CC Bacteria; Cyanobacteria; Chroococcales; Synecchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Murauchi C., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Matanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: D90910; BA17944.1; -.
CC DR HSSP: Q9SWE5; IE20.
CC DR InterPro: IPR005252; Cons_hypoth521.
CC DR InterPro: IPR003382; Flavoprotein.
CC DR Pfam: PF02441; Flavoprotein; 1.
CC DR TIGRFAMS: TIGR00521; dfp; 1.
CC KW Flavoprotein; Complete proteome.
CC SO SEQUENCE 402 AA: 43419 MW; 7E3F73DE185EC97C CRC64;
CC -----
Query Match 20.9%; Score 239.5; DB 1; Length 402;
Best Local Similarity 37.3%; Pred. No. 7.7e-16;
Matches 66; Conservative 25; Mismatches 69; Indels 17; Gaps 6;
OY 29 RLLAASGSVASIKESN-LCHCFSEWAEVKAASKSLNFWKPSLPQWNT-----LYT 81
DB 6 RLILGVGGIAAKICEVVSOLFQGAEVRYLTAEKRV-----TPLFTTLARPAYG 61
OY 82 DEDEWSSNKKIGDPYLHIELRRWADVMIAPLSANTLAKIAGLGDNLITCIVRAMDYSK 141
DB 62 DAD--FWQPIHHRPLHLDIGEMADIFLAPLHATLAKLGHFADLLSNVLA--SSG 116
OY 142 PLFVAPAMNTLMNPNTERHL-VLDELGITLIPPIKKRLACGDYGCNMAEPLSI 197
DB 117 PLLAPAMNTDMWEDBAVQNRNLQQLIDRRYHLAPNGGLACDRGVRLAEPAOI 173
CC -----
RESULT 7
DFF_HAEIN STANDARD: PRT; 400 AA.
AC P44953;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA/pantothenate metabolism Flavoprotein homolog.
GN DFP OR HI0953.
OS Haemophilus influenzae.
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrichback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC -----
DR EMBL: U32776; AAC22614.1; -
DR HSSP: Q9SWE5; IE20.
DR TIGR: H10953; -
DR InterPro: IPR005252; Cons_hypoth521.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein.1.
DR TIGRfams: TIGR00521; dfp.1.
DR Flavoprotein: Complete proteome.
KW SEQUENCE 400 AA; 43401 MW; 1D6FEAE20C26B913 CRC64;

Query Match 20.7%; Score 236.5; DB 1; Length 400;
Best Local Similarity 32.3%; Pred. No. 1.5e-15;
Matches 61; Conservative 38; Mismatches 67; Indels 23; Gaps 4;

QY 30 ILLAASGSVASIKFSNLCHESE-MAEVKAVASKSLNFVDKPSLPQNTLYTDEDEMS 88
DB 8 IYVGITGGIAAKYKTELIRLLKAEAEVAVVLTAPAAAEVTPLT-----QAIS 56
QY 89 WNKIDPVL-----HIELRRADVMIAPLSANTLAKIAGLCNLTCTIRAMDY 139
DB 57 GNAVSQSLDPOAELAMGHIELAKWADAIILAPASADFIARLTIGMANDLSTICLA--T 114
QY 140 SKPLVAPAMNTLMNNPTEERHVLVDDELGITLPIPKKACGYGGAAPSLIYS 199
DB 115 NAPIFLAPAMNQMHQSTIQONLTTLQTRGIELIGPNSGFOACGDMGRMSEPEIFT 174
QY 200 TVRLFWEQ 208
DB 175 ALSDFFSOK 183

RESULT 8
DFP_STRMU STANDARD; PRT: 145 AA.
AC 054433:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein homolog (Fragment).
GN DFP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG8.
RX MEDLINE=96312359; PubMed=8763945;
RA Gutierrez J.A., Crowley P.J., Brown D.P., Hillman J.D.,
RA Youngman P., Bleiweis A.S.;
RA "Insertional mutagenesis and recovery of interrupted genes of
RT Streptococcus mutans by using transposon Tn917: preliminary
RT characterization of mutants displaying acid sensitivity and
RT nutritional requirements.";
RL J. Bacteriol. 178:4166-4175(1996).
CC -I- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U48885; AAC44502.1; -
CC HSSP: Q9SWE5; IE20.
CC InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein.1.
DR Flavoprotein: Complete proteome.
KW SEQUENCE 400 AA; 43401 MW; 1D6FEAE20C26B913 CRC64;
FT NON_TER 1 1

```

```

FT NON_TER 145 145
SQ SEQUENCE 145 AA; 15819 MW; 5920CDD5708F2DAB CRC64;

Query Match 19.0%; Score 217.5; DB 1; Length 145;
Best Local Similarity 35.2%; Pred. No. 3e-14;
Matches 51; Conservative 30; Mismatches 57; Indels 7; Gaps 3;

QY 29 ILLAASGSVASIKFSNLCHESEMA-EVKAVASKSLNFVDKPSLPQNTLYTDEDEMS 87
DB 4 KILLAVSGSIAAYKADLSHQTLKGYHNVFWMTNAKQFIP----PLTQVLSKNPVYS 59
QY 88 SNWKIDP--VLHIELRRADVMIAPLSANTLAKIAGLCNLTCTIRAMDYSPFLV 145
DB 60 NWKEDDPQVINIHALAKADFLPFPASANTLAHLHGFDADIVTSVALAPLEVPKFF 119
QY 146 APAMNTLMNNPTEERHVLVDDELG 170
DB 120 APAMNTKMTENPTQSNITLLKKFG 144

RESULT 9
DFP_ECOLI STANDARD; PRT: 406 AA.
AC P24285; P76718;
DT 01-MAR-1992 (Rel. 21, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein.
GN DFP OR B3639.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
RN [2]
RP SEQUENCE OF 288-406 FROM N.A.
RC STRAIN=KL2;
RX MEDLINE=84057777; PubMed=6139280;
RA Lundberg L.G., Thoresson H.O., Karlstrom O.H., Nyman P.O.;
RT "Nucleotide sequence of the structural gene for dnfPase of
RT Escherichia coli K-12.";
RL EMBL J. 2:967-971(1983).
CC -I- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: F10328; AAA61992.1; ALT_INIT.
CC EMBL: AE000441; AAC76663.1; ALT_INIT.
CC EMBL: V01578; -; NOT_ANNOTATED_CDS.
CC HSSP: Q9SWE5; IE20.
CC Ecogene: E610004; dfp.
CC InterPro: IPR005252; Cons_hypoth521.
CC InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein.1.
DR TIGRfams: TIGR00521; dfp.1.
DR Flavoprotein: Complete proteome.
KW SEQUENCE 406 AA; 43438 MW; CBD11B9347E8C6AB CRC64;

Query Match 17.9%; Score 205.5; DB 1; Length 406;

```



```

Db 111 A--STAPYFALAMNMWYNPLIKENIKLTKTYNKYFIEPPDKGLACSSNALGRKLNED 168
OY 196 LIYSTV 201
Db 169 KIKIT 174

RESULT 12
DFF_MYCTU STANDARD: PRT; 418 AA.
AC P71661;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA/pantothemate metabolism flavoprotein homolog.
GN DFP OR RV1391 OR MT1436 OR MTCY2184.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishel W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
CC -i- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL: Z80108; CAB02174.1; -
CC DR EMBL: AE007015; AAK45701.1; -
CC DR HSSP: O9SWE5; IE20.
CC DR TIGR: MT1436; -
CC DR Tuberclust; RV1391; -
CC DR InterPro: IPR005252; Cons_hypoth521.
CC DR InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein; 1.
CC DR TIGRFAMs: TIGR00521; dff; 1.
CC KW Flavoprotein; Complete proteome.
CC SEQUENCE 418 AA; 43577 MW; 48BD95E536595506 CRC64;
Query Match 15.4%; Score 176.5; DB 1; Length 418;
Best Local Similarity 24.7%; Pred. No. 1,le-09;
Matches 49; Conservative 38; Mismatches 66; Indels 45; Gaps 4;

```

```

OY 29 RILLAAGSVASIKFSNLCCHPSEMA-EVKAVASKSLNPFV----- 68
Db 10 QVIVGSGGIIAAKACTVYRQULEASHRVVITPESALRFRVCAATFEALSCEPVCTDVEA 69
OY 69 DKPSLPONTVLTDEDEWESWKNIGDPVLHIELRRADVMIAPLSANTLAKTAGLGDN 128
Db 70 DVPAVP-----HVHIGQADLVVAPATADLLARAAGRAD 106
OY 129 LITCIVRANDYSKPLFVAPAMNTLMNPNPTERRHVLDELGITLIPRIKKIACGDYGN 188
Db 107 LITATLTLT--ARCPVLEFADAMHTEMLHPATVDNVATLRRGAVVLEPATGRLTGADSGA 164
OY 189 GAMAEPSLIYSTVRLFE 206
Db 165 GRUPEAEIITTLAQLLE 182

RESULT 13
DFF_METUA STANDARD: PRT; 403 AA.
AC Q58323;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA/pantothemate metabolism flavoprotein homolog.
GN Mj0913.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxId=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8668087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleink H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -i- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL: U67535; AAB98918.1; -
CC DR TIGR: M00913; -
CC DR InterPro: IPR005252; Cons_hypoth521.
CC DR InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein; 1.
CC DR TIGRFAMs: TIGR00521; dff; 1.
CC KW Hypothetical protein; Flavoprotein; Complete proteome.
CC SEQUENCE 403 AA; 45670 MW; B8B51AA81A115B94 CRC64;
Query Match 11.8%; Score 135.5; DB 1; Length 403;
Best Local Similarity 27.1%; Pred. No. 1,le-05;
Matches 49; Conservative 34; Mismatches 61; Indels 37; Gaps 7;
OY 3 GTSSPLGCKKIKKKMMEDVTYRKPRILLAAGSVASIKFSNLC-HCPSEMAEYKAVAS 61
Db 14 GTSKILENK-----KITVAVTSSIAIALETFRKLARELIRHGAEEYCIIT 57

```



```
OY 62 KSSLNFVDPKPSLPQNVTLTYTDEDEWSSMKN1-----GDPVLHIELRRMADVMIAPLSAN 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 EETKRIIGREAL-----KFGCGNEVEYEITGD-IEHILTYNECCCLLIYPATAN 105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 117 TLAKIAGGLCDNLITICIVAMDYSKPLFVAPAMNTLMNNPFTERRHLVLLDELG-ITLIP 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 106 IISKINLGIADINIVNTTALMFQGNKPIFTVPAMHENMFN--AIKRHIIDIKLEKDKYIIS 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 176 P 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 164 P 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
PAAD_MENJA STANDARD; PRT; 184 AA.
AC 057566:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN MJ0102.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAI-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerelevage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overback R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Urtreback T.R., Kelley J.M., Peterson J.D., Sadow F.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RT Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U67467; AAB98082.1; -
CC TIGR: MJ0102;
DR InterPro: IPR003382; Flavoprotein.
DR InterPro: IPR004507; Ubix.
DR Pfam: PF02441; Flavoprotein; 1.
DR TIGRFAMs: TIGR00421; ubix; 1.
KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 184 AA; 20643 MW; F3EDBF5B48A82C5 CRC64;

Query Match
Best Local Similarity 9.5%; Score 109; DB 1; Length 184;
Matches 42; Conservative 24; Mismatches 61; Indels 40; Gaps 7;
```

```
OY 138 DVS-----KPLFVAPAMNTLMNNPFTERRHL---VLLDELGITLIPPI 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 106 DIALERRKLIIMPR-----EXPFNSHLENNLKLSNIGALVMPPI 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
PAAD_SYNY3 STANDARD; PRT; 206 AA.
AC P22743;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN SLR1099.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria: Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RT DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: D90900; BAA16758.1; -
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 206 AA; 22172 MW; 7C1E13F47996E3F CRC64;

Query Match
Best Local Similarity 9.4%; Score 108; DB 1; Length 206;
Matches 48; Conservative 24; Mismatches 56; Indels 52; Gaps 10;

OY 30 ILAASG-----SVASIKFSNLCHESEMAEYKAVASKSLNF-----VDKPSLPQNV 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 LILGVGASGLIYAVRAIK-----HLLADYITIELVYASRSYVWQAEQNIQGPES-- 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 78 TLYTDEDEWSSW--SKIGDPVLHIELRRMADV-----MIAPLSANTLANI 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 ---AQAEFMSQAGVEKGLKICH---RMGDVGATIASGSRGACGMVYLPCSMSTYAKL 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 122 AGGLCDNLITCIYRANDY-----SKPLFVAPAMNTLMNNPFTERRHLVLLDELGITLIPPI 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 AVGMSSDLLE---RAADVQIKEGKPLVVVPRETP--SLIHRLNLTSLAEAGVRIYPAI 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: November 25, 2002, 10:14:53
Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 10:13:41 ; Search time 21 Seconds

(Without alignments)
993.389 Million cell updates/sec

Title: US-09-868-300-8

Perfect score: 1145
Sequence: 1 EFGTSFLGCKIKKKMMME.....YSTVRLFWESQARKQDQTS 217

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	421.5	36.8	625	2 T37719	thymidylate syntha
2	421	36.8	237	2 T21376	hypothetical prote
3	347	30.3	571	2 S37913	hypothetical prote
4	331.5	29.0	562	2 S58149	S152 protein - yea
5	324.5	28.3	531	2 S57752	hal3 protein - yea
6	323	28.2	674	2 S66937	probable membrane
7	280.5	24.5	388	2 S60371	pantothenate metab
8	270.5	23.6	410	2 A82194	pantothenate metab
9	264.5	23.1	183	2 F98010	pantothenate metab
10	264.5	23.1	183	2 H95142	N-terminal region
11	256.5	22.4	412	2 F87709	flavoprotein limpo
12	249.5	21.8	394	2 A81058	DNA/pantothenate m
13	246.5	21.5	394	2 DB1819	DNA/pantothenate m
14	243.5	21.3	178	2 DB6696	hypothetical prote
15	240	21.0	404	2 A13538	pantothenate metab
16	239.5	20.9	402	2 S75082	phosphopantocheny
17	239.5	20.9	404	2 F83963	pantothenate metab
18	237.5	20.7	398	2 C97112	flavoprotein dfr f
19	236.5	20.7	400	2 G64104	flavoprotein invol
20	232.5	20.3	399	2 F89893	pantothenate metab
21	228.5	20.0	399	2 A11674	hypothetical prote
22	227.5	19.9	399	2 A11302	pantothenate metab
23	226.5	19.8	405	2 F81354	pantothenate metab
24	225	19.7	405	2 A10006	DNA/pantothenate
25	224.5	19.6	406	2 D69878	pantothenate metab
26	224	19.6	425	2 E71889	pantothenate metab
27	223.5	19.5	402	2 B82962	DNA/pantothenate m
28	221	19.3	425	2 A64625	pantothenate metab
29	216.5	18.9	399	2 A82351	DNA/pantothenate m

30	210.5	18.4	394	2 A72223	pantothenate metab
31	210.5	18.4	409	2 H82842	DNA/pantothenate m
32	207.5	18.1	407	2 AC0971	conserved hypothet
33	205.5	17.9	430	1 A65165	pantothenate metab
34	205.5	17.9	430	2 B91193	pantothenate metab
35	205.5	17.9	430	2 C66040	pantothenate metab
36	204.5	17.9	181	2 S23418	pantothenate metab
37	204	17.8	401	2 AD2615	lactibiotic epider
38	184	16.1	390	2 C70201	pantothenate metab
39	183.5	16.0	185	2 B95019	pantothenate metab
40	176.5	15.4	367	2 C97397	flavoprotein limpo
41	176.5	15.4	418	2 E70899	pantothenate metab
42	173.5	15.2	419	2 G66976	pantothenate metab
43	169	14.8	401	2 E75114	probable flavoprot
44	168.5	14.7	413	2 B90162	DNA/pantothenate m
45	167.5	14.6	401	2 G71018	DNA/pantothenate m

ALIGNMENTS

```
RESULT 1
T37719
thymidylate synthase-like DNA metabolism protein - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37719
R:Murphy, L., Harris, D., Barrell, B.G., Rajandream, M.A., Lyne, M.H.
Submitted to the EMBL Data Library, August 1999
A:Reference number: Z21739
A:Accession: T37719
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-625 <MUP>
A:Cross-references: EMBL:AL109770; PIDN:CB52423.1; GSPDB:GN00066; SPDB:SPAC15E1.04
A:Experimental source: strain 97Zh-; cosmid c15E1
C:Genetics:
A:Gene: SPDB:SPAC15E1.04
A:Map position: 1

Query Match      36.8%; Score 421.5; DB 2; Length 625;
Best Local Similarity 48.8%; Pred. No. 6.1e-33;
Matches 84; Conservative 31; Mismatches 52; Indels 5; Gaps 3;

QY 27 KPRILLASGVASIRKFSNLCHCFSEW--AEYKAVASKSLFVDPKPSLPQ--NTLTLYDE 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 30 KHIILVAATGSAATKILIKLYKSLTTYGVQVYVLTDPARNFVEKEDITLGVAVYNNNA 89
QY 84 DEMSSMNRIGDPVLEHIEELRRWADVMIIAPLSANTLAKTAGICDNLITCIYRWDYSKPL 143
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 90 DDMKRWMDGIECPITHTIELRRMHLLILAPLSANTMAKMANGLCDLILSLIRAMPPLKPI 149
QY 144 EVAPAMNTLMNNPFTENHLVLDL--GTLIPPIKKLACGDYGAMAE 193
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 150 ILAPAMNTLMNTNPITOEHLAISRIYKNSFIMPIERVLACGDIGMGMAE 201

RESULT 2
T21376
hypothetical protein F25H9.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21376
R:Smey, R.
Submitted to the EMBL Data Library, October 1996
A:Reference number: Z19414
A:Accession: T21376
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-237 <WII>
A:Cross-references: EMBL:Z81069; PIDN:CA802993.1; GSPDB:GN00023; CESP:F25H9.6
A:Experimental source: clone F25H9
C:Genetics:
```


QY 164 YLL--DELGITLPIPIKKKL-ACGDYNGAMAE 193
DB 414 RLADMPWTEVLKPLEKVGSTGDIKMGKMTD 446

RESULT 6
S66937
probable membrane protein YOR054c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein O2801
C:Species: *Saccharomyces cerevisiae*
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C:Accession: S66937; S66928
R:Bohr, C.; Bolotin-Fukuhara, M.; Baigian-Fornier, B.; Dang, D.V.; Valens, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66929
A:Accession: S66937
A:Molecule type: DNA
A:Residues: 1-674 <BOH>
A:Cross-references: EMBL:Z74961; NID:q1420188; PID:e251974; PID:q1420190; MIPS:YOR054c
A:Experimental source: strain S288C
R:Landt, O.; Hiesel, R.; Unseld, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66907
A:Accession: S66928
A:Molecule type: DNA
A:Residues: 674 <LAN>
A:Cross-references: EMBL:Z74961; MIPS:YOR054c
A:Experimental source: strain S288C
C:Genetics:
A:Cross-references: SGD:S0005580
A:Map position: 15R
C:Keywords: Transmembrane protein
F:468-484/Domain: transmembrane #status predicted <TM>

Query Match 28.2%; Score 323; DB 2; Length 674;
Best Local Similarity 36.6%; Pred. No. 2.7e-23;
Matches 71; Conservative 40; Mismatches 57; Indels 26; Gaps 5;

QY 11 NKEKKMNEVDVTYRPRLLAASGSVASIKFSNLCHCFSEMAEYKAVASKSLNFVDR 70
DB 395 NQOEKQNA--STGCFPTL-----GGSRTYSNSNVVSGHPQIE----- 433

QY 71 PSLPQNVTLTYTDEDESSNNKIGDPLVLIHLRRMADVMITAPLSANTLAKIAGGLCDNLL 130
DB 434 -LPALIDQWTDQEDWVRORTDPLVLIHLRRMADILVAPLANTLAKIALGLCDNLL 491

QY 131 TCIVRAMDSKPLFVAPAMNTLMNNPTEERHLVYLDE--LGTILPIPIKKKLAC-GDYG 187
DB 492 TVYIRAMNTEPFIPLAFPSKSGCTFNSIMTKHFRITIQEPMVTVKPSKVGWINGDIG 551

QY 188 NGAMAEPLIYSTV 201
DB 552 LSGMDANEIVGKI 565

RESULT 7
B70371
pantothenate metabolism flavoprotein dfp - *Aquifex aeolicus*
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: *Aquifex aeolicus*
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 29-Sep-1999
C:Accession: B70371
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oy
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70371
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-388 <AO>
A:Cross-references: GB:AE000708; NID:g2983356; PIDN:AAC06944.1; PID:g2983357; GB:AE00065
A:Experimental source: strain VF5

C:Genetics:
A:Gene: dfp
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 24.5%; Score 280.5; DB 2; Length 388;
Best Local Similarity 41.2%; Pred. No. 1.9e-19;
Matches 70; Conservative 26; Mismatches 55; Indels 19; Gaps 7;

QY 30 ILLAASGSVASIKFSNLCHCFSEMA---EYKAVASKSLNFVDRKPSLPQNT--LYTD 82
DB 3 ILIGVGGIASYK---VCELVRELKRGKSHSVKTLTPFAKEFW-SPLTFQTLISGNKAYTD 58

QY 83 EDEWSSWNKIGDPLVLIHLRRMADVMITAPLSANTLAKIAGGLCDNLLTCIVRAMDSKP 142
DB 59 KD---WEE--EPLAHINLARWADVFLIAPATANTLAKIANGISDMLLTTLIA--YGRP 110

QY 143 LPVAPAMNTLMNNPTEERHLVYLDELTITLPIPIKKKLACGDYNGAMA 192
DB 111 LTVAPAMNTVMKSPSTQENLKLKEMGHVITIEPFGVLACKEVGGKLA 160

RESULT 8
AH2194
pantothenate metabolism flavoprotein [imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AH2194
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matsumoto, A.; Irigun
Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074810.1; PID:q17132205; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alh111
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 23.6%; Score 270.5; DB 2; Length 410;
Best Local Similarity 38.0%; Pred. No. 1.9e-18;
Matches 76; Conservative 29; Mismatches 74; Indels 21; Gaps 7;

QY 26 KRPRILLAASGSVASIK---FSNLCHCFSEMAEYKAVASKSLNFVDRKPSLPQNTL-- 79
DB 11 KRSRYLVAVGGGIAAYKCEVVSNTL---FKSGVEVAVILTRSAOEFITPLTSL--TLSR 64

QY 80 ---YTDEDESSNNKIGDPLVLIHLRRMADVMITAPLSANTLAKIAGGLCDNLLTCIVRA 136
DB 65 HPAYTDDDFWQSTH---SRPLHIELEGWADLVIAPLRANTLAKLTYGADNLLTNTVLA 121

QY 137 WDYSKPLFVAPAMNTLMNNPTEERH-LVLDDELGTITLPIPIKKKLACGDYNGAMAEPS 195
DB 122 --SNCPALLAPAMNTDMWQLTVGRWMOQLTDSRYHGIGTASGLLACDRVAGARNAEPP 179

QY 196 LIYSIVRLFWESQARKQRDG 215
DB 180 ETLAYTQSLHTHQGRDLAG 199

RESULT 9
F98010
N-terminal region of Dfp protein homolog [imported] - *Streptococcus pneumoniae* (strain
C:Species: *Streptococcus pneumoniae*
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 03-Jun-2002
C:Accession: F98010
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genom, Y. of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; M0ID:21429245; PMID:11544234
 A:Accession: F98010
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-183 <KUR>
 A:Cross-References: GB:AE007317; PIDN:AAK99914.1; PID:915458737; GSPDB:GN00174
 C:Genetics:
 A:Gene: dfp N-terminus
 C:Superfamily: conserved hypothetical protein M0730

Query Match 23.1%; Score 264.5; DB 2; Length 183;
 Best Local Similarity 36.0%; Pred. No. 2.7e-18;
 Matches 64; Conservative 39; Mismatches 64; Indels 11; Gaps 4;

Oy 30 ILLAASGVASIFSNCHCFSEWA-EVKAVASKSLNFPVKPS--LPON-VTLTYDDED 84
 Db 4 ILLAVTGSTIASYSADLVSSLSKKGHOVTVLMTQAATFPIQLTVLSQNPVHLDMKE 63
 Oy 85 EWSNMKIGDPVHLIELRMADVMIIPASNTLAKIAGGLCDNLTICIVRAMDYSKPLF 144
 Db 64 PYP-----DQVNHIELGKRADELFTVVPATANTIAKLHAGFADNVTSTALALPSHIIRL 117
 Oy 145 VAPAMTLMKNNPETERHVLVLDLDELITLIPPIKKIACGDYGANGAMAEPSLIYSTVR 202
 Db 118 IAPAMTKMYDHPVTONNLTLETTYGQLIAPKESILACGDHGRGALADLTIIILERIK 175

RESULT 10
 H95142
 Flavoprotein [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Jun-2002
 C:Accession: H95142
 R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; M0ID:21357209; PMID:11463916
 A:Accession: H95142
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-183 <KUR>
 A:Cross-References: GB:AE005672; PIDN:AAK75337.1; PID:914972713; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SPI231
 C:Superfamily: conserved hypothetical protein M0730

Query Match 23.1%; Score 264.5; DB 2; Length 183;
 Best Local Similarity 36.0%; Pred. No. 2.7e-18;
 Matches 64; Conservative 39; Mismatches 64; Indels 11; Gaps 4;

Oy 30 ILLAASGVASIFSNCHCFSEWA-EVKAVASKSLNFPVKPS--LPON-VTLTYDDED 84
 Db 4 ILLAVTGSTIASYSADLVSSLSKKGHOVTVLMTQAATFPIQLTVLSQNPVHLDMKE 63
 Oy 85 EWSNMKIGDPVHLIELRMADVMIIPASNTLAKIAGGLCDNLTICIVRAMDYSKPLF 144
 Db 64 PYP-----DQVNHIELGKRADELFTVVPATANTIAKLHAGFADNVTSTALALPSHIIRL 117
 Oy 145 VAPAMTLMKNNPETERHVLVLDLDELITLIPPIKKIACGDYGANGAMAEPSLIYSTVR 202
 Db 118 IAPAMTKMYDHPVTONNLTLETTYGQLIAPKESILACGDHGRGALADLTIIILERIK 175

RESULT 11
 F87709
 DNA/pantothenate metabolism flavoprotein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

A:Accession: F87709
R.Nierman, W.C.; Feldbljum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MUID:21173698; PMID:11259647

A:Accession: F87709
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-412 <STO>
A:Cross-references: GB:AEO05673; NID:g13425480; PIDN:AAK25674.1; GSFPDB:GN00148

A:Gene: CC3712
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 22.4%; Score 256.5; DB 2; Length 412;
Best Local Similarity 32.1%, Pred. No. 4,Se=17/
Matches 63; Conservative 39; Mismatches 73; Indels 21; Gaps 4;

Oy 29 RLILASGSVASIKFSNLCHCFSEMA-WAEVKAVAKSSLNPFVDKPSP-----QNVTL 79
| : | : || | : | : | : | : | : | : | : | : | : | :
Db 19 RVLLVGGGVAAYKKALLTRLRKGAVRPRLTKAGAAFPTPLSLALADKYEDLFS 78

Oy 80 YTDDEWSSWNKIGDPVLHIELRWADVMIAPLSANTLARISAGLCDNLTCTIVRANDY 139
||| | | : | : | : | : | : | : | : | : | : | : | :
Db 79 LTDERHWG-----HIELRSADLVVVAPAATADIARIAAQGLADLASTTLLARD 127

Oy 140 SKPFLVARAMNTLMNNNPETERHLVLDDELGITLIPIPKKLACGDYGNGAMPESLIYS 199
|| : |||| | : || : | : | : | : | : | : | : | : | : | :
Db 128 -kPVLAAPMANVRMLHPATORNIATLEADGVRFVPPEGAMAGEFGPGRIABEAIFA 186

Oy 200 TVRLFWSQAQRKRQG 215
: | : | : | : | : | : | : | : | : | : | : | : | :
Db 187 ALMALLEGPARPLEG 202

RESULT 12
AB1058
DNA/pantothenate metabolism flavoprotein NMB1658 [Imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

R:Accession: AB1058
R:Tettelein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.,
Tri, H.; Qin, H.; Yamaherap, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000

A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunoli, R.:
A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307

A:Accession: AB1058
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <TEXT>
A:Cross-references: GH:AEO02516; GB:AEO02098; NID:g7226905; PIDN:AAF42007.1; PID:g7222

A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1658
C:Superfamily: pantothenate metabolism flavoprotein.dfp

Query Match 21.8%; Score 249.5; DB 2; Length 394;
Best Local Similarity 35.7%, Pred. No. 2e-16;
Matches 70; Conservative 24; Mismatches 71; Indels 31; Gaps 5;

Oy 30 ILILAASGSVASIKFNSNLCHFSEMA-EVKAVAKSSLNFVDKPSLPQNVTLYTDEDEMSS 88
||| :|| : | : | : | : | : | : | : | : | : | : | : | :
Db 5 LLVGTVGISIAAKSCELVRLKKGSHSVTVVMRSRAPEVF-SPLTFQLKS----- 53

Oy 89 WKKTDDPVL-----HTLRWDADMVIITAPSANTLAKIAGCDCLDLTCIVR 135
||| :||| | : | : | : | : | : | : | : | : | : | : | :
Db 54 -----GNPVLDTHGNGSGMGEINILTNRNADFVLTAANSMRTVAKICGTGVDNDNLTSIAA 109

THIS PAGE BLANK (USPTO)

Gencore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using sw model

Run on: November 25, 2002, 10:15:36 ; Search time 10 Seconds
(without alignments)
339.854 Million cell updates/sec

Title: US-09-868-300-8
1145
1 EFGTSFLGCKIKKMKME.....YSTVRLFWESQARKORDGTS 217

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	422	36.9	170	10	US-09-764-860-585
2	221	19.3	228	9	US-09-895-913A-66
3	188.5	16.5	188	12	US-10-047-676A-9
4	76.5	6.7	130	10	US-09-771-383-13
5	73.5	6.4	372	10	US-09-873-880-40
6	73.5	6.4	1395	9	US-09-808-602-67
7	72	6.3	426	12	US-10-036-568-2
8	71.5	6.2	780	9	US-09-935-868-34
9	71.5	6.2	780	9	US-09-935-868-38
10	71.5	6.2	780	9	US-09-935-868-42
11	71.5	6.2	782	9	US-09-935-868-48
12	71.5	6.2	782	9	US-09-935-868-52
13	70.5	6.2	2441	12	US-10-109-886-8
14	70.5	6.2	2442	12	US-10-109-886-10
15	70	6.1	194	10	US-09-870-162A-9
16	70	6.1	264	10	US-09-815-242-11559
17	69.5	6.1	394	10	US-09-993-844-7
18	69.5	6.1	408	10	US-09-895-211-5
19	69	6.0	90	10	US-09-354-453-52

20	69	6.0	236	10	US-09-803-286A-5	Sequence 5, Appli
21	69	6.0	429	10	US-09-867-852-109	Sequence 109, App
22	69	6.0	434	10	US-09-921-823-20	Sequence 20, Appl
23	69	6.0	462	10	US-09-846-729A-3	Sequence 3, Appli
24	69	6.0	462	10	US-09-846-729A-17	Sequence 17, Appl
25	69	6.0	639	10	US-09-835-996A-15	Sequence 15, Appl
26	68.5	6.0	408	10	US-09-895-211-2	Sequence 2, Appli
27	68	5.9	199	10	US-09-764-864-1228	Sequence 128, Ap
28	68	5.9	718	10	US-09-801-368-388	Sequence 388, App
29	67	5.9	793	9	US-09-935-868-32	Sequence 32, Appl
30	67	5.9	1464	10	US-09-746-390-2	Sequence 2, Appli
31	66.5	5.8	1025	10	US-09-854-886-2	Sequence 2, Appli
32	66	5.8	189	10	US-09-815-242-10229	Sequence 10229, A
33	66	5.8	430	10	US-09-815-242-13726	Sequence 13726, A
34	66	5.8	439	10	US-09-815-242-13917	Sequence 13917, A
35	66	5.8	495	10	US-09-925-301-1312	Sequence 1312, Ap
36	65.5	5.7	96	10	US-09-283-059-1208	Sequence 1208, Ap
37	65.5	5.7	382	10	US-09-547-267-7	Sequence 7, Appli
38	65.5	5.7	590	9	US-10-115-701A-2	Sequence 2, Appli
39	65	5.7	324	10	US-09-841-132-496	Sequence 496, App
40	65	5.7	393	10	US-09-815-242-12395	Sequence 12395, A
41	64.5	5.6	80	10	US-09-864-761-37611	Sequence 37611, A
42	64.5	5.6	110	10	US-09-283-059-358	Sequence 358, App
43	64.5	5.6	317	12	US-10-006-867-116	Sequence 116, App
44	64.5	5.6	923	12	US-10-104-440-2	Sequence 2, Appli
45	64	5.6	245	10	US-09-911-826A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-764-860-585
Sequence 585, Application US/09764860
Patent No. US20020094953A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC008
CURRENT APPLICATION NUMBER: US/09/764,860
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1198
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 585
LENGTH: 170
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-860-585

Query Match 36.9%; Score 422; DB 10; Length 170;
Best Local Similarity 52.3%; Pred. No. 2e-39;
Matches 81; Conservative 20; Mismatches 50; Indels 4; Gaps 2;

QY	55	EYKAVASKSLNFVDKPSLPQWVLTYTEDESSWNKIGDPVLIETLRMADVMTIAPLS	114
DB	13	EVAVVTTERAKHFYSPQDIP--VTLYSDADEWEMKSSSDPVLHIDLRKMDLLVAPLD	70
QY	115	ANTLAKIAGGLCDNLTCTVRAWYDYSKPLFVAPAMNTITLMNNPFERHLYLDELGITILI	174
DB	71	ANTLAKVAGSGIDNLTCTVRAWYDYSKPLFVAPAMNTITLMNNPFERHLYLDELGITILI	130
QY	175	PIPKKLAGDYGNGAMAEPSLIYSTVR--LFWES	207
DB	131	PCVAKKLVCGDGLGAMAEVGTIVDKVKEVLFQHS	165
RESULT 2			
US-09-895-913A-66			
Sequence 66, Application US/09895913A			
Patent No. US20020160456A1			
GENERAL INFORMATION:			
APPLICANT: Kleantous, Harold			


```

QY 68 VDRPSL-----PQNY-----77
      : |
Db 92 YENGIGTIGCVHRTYRNNDGTMIDIKYVAIRHPDGALYYPTTLLCLENTHANCGG 151
      : |
QY 78 ----TLVTDDEWSSWMKIGDPV-----LHIELRMDAWIIAPLSANTLAKIAGGLCD 127
      |||| :||| :| :| :| :| :|
Db 152 KCLSAEYTD-----VGEVAKSHGLKHLIIDARLFNMSVALGCVHRLKYAA----D 199
      :||| :||| :| :| :| :| :|
QY 128 NLITCIVRAMDYSKPLFVAPAMNTLMNPFTEHVLIVBELG 170
      :| :| :| :| :| :| :| :|
Db 200 SVSVCT-----SKGL-GAPYGVSIYGVSTAFIEAKITLKRITLG 235

```

```

RESULT 6
US-09-808-602-67
; Sequence 67, Application US/09808602
; Patent No. US2002015115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kunud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NO. US2002015115A1 Proteins and Nuclec Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-808-602-67

```

```

Query Match          6.4%; Score 73.5; DB 9; Length 1395;
Best Local Similarity 27.6%; Pred. No. 14;
Matches 34; Conservative 16; Mismatches 44; Indels 29; Gaps 7;

OY      50  FSEMAEYKAVASKSLNFVDPKSLPON--VTLTYDEDEWSMKNKIGDVLH----- 98
          | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db      746  FFETIEGOPSNSKTALTYEDVPSAPPDNIQIGMYNQTAGVWWTLP--PPSQHHNGNLXGY 803

OY      99  -IELRRMADVMIIAPLSANLTAKINGGLCDNLFFCTV-----RAND--YSKP--LF 144
          || : : : : : | : : : : | : : : : | : : : : | : : : : |
Db      804  KIEVSAGFTMKVLANMTLN--ATTTSVLLNLTGAVYSVRLNSTFTKAGDGPYSKPSLSE 861

OY      145  VAP 147
          : |
Db      862  MDP 864

RESULT 7
US-10-036-568-2
; Sequence 2, Application US/10036568
; Patent No. US20020090682A1
; GENERAL INFORMATION:
; APPLICANT: Willson, Tracy
; APPLICANT: Nicola, Nicos A.
; APPLICANT: Hilton, Douglas J.
; APPLICANT: Metcalf, Donald
; APPLICANT: Zhang, Jian G.
; TITLE OF INVENTION: NOVEL HAEMOPOIETIN RECEPTOR AND GENETIC SEQUENCES
; TITLE OF INVENTION: ENCODING SAME
; FILE REFERENCE: Davies cc
; CURRENT APPLICATION NUMBER: US/10/036,568

```

```

? CURRENT FILING DATE: 2001-11-07
? PRIOR APPLICATION NUMBER: US/09/051,843
? PRIOR FILING DATE: 1998-06-29
? NUMBER OF SEQ ID NOS: 11
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 2
? LENGTH: 426
? TYPE: PRT
? ORGANISM: nuc. & predicted a.a. seq. of mnr4
? FEATURE:
? NAME/KEY: unsure
? LOCATION: (12)
? OTHER INFORMATION: authors are unsure about the sequence assignment
? NAME/KEY: unsure
? LOCATION: (194)
? OTHER INFORMATION: authors are unsure about the sequence assignment
? IS-10-036-568-2
```

	Query Match	6.3%;	Score 72;	DB 12;	Length 426;	
	Best Local Similarity	19.3%;	Pred. No. 4;			
	Matches	43;	Conservative	23;	Mismatches	49;
					Indels	108;
					Gaps	9;
<hr/>						
OY	10	CNKIEKKKKNMEVDIYTRKPRILLAAAGSY-ASIKRSNLC----	HCFSEKAELVAKAASKS	64		
		: : : : : : : : : :				
Db	281	CONSIEDNRMKETSCFOGLPVADVVYTVRAVRANKLCEFDNNKLMSDSMAOSIGKE--		338		
<hr/>						
OY	65	LNFVDKPSPONVTLYTPDEDEWSSNNKIGDPYLHIELRRADVMIAPLSANTLAKIAGG		124		
Db	339	-----ONSTFYT-----		345		
<hr/>						
OY	125	LCDNILTCIVRAMDYSKPLFVAAPAMNTLMNNPFTERHLVLLDELGITLIPPIKKRLACG		184		
		: : : : : : : : : :				
Db	346	-TMLLT-----EVFAVAVIIIL-----LFYLRKLIITIFPPI-----		377		
<hr/>						
OY	185	DYGCAAMAEPISLYSTV-----RLFV-----BSQARKQMD		214		
		: : : : : : : : : :				
Db	378	-----DDPGRIEKFEMFGDONDTLLMHKKDIYERKOSKEED		413		

```

: RESULT 8
: US-09-935-868-34
: Sequence 34, Application US/09935868
: Patent No. US20020164690A1
: GENERAL INFORMATION:
: APPLICANT: Regeneron Pharmaceuticals, Inc
: TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using
: FILE REFERENCE: REG 203D
: CURRENT APPLICATION NUMBER: US/09/935,868
: CURRENT FILING DATE: 2002-04-11
: PRIOR APPLICATION NUMBER: PCT/US99/22045
: PRIOR FILING DATE: 1999-09-22
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 34
: LENGTH: 780
: TYPE: PRT
: ORGANISM: Homo Sapiens
: US-09-935-868-34

```

	Query March	6.2%	Score 71.5;	D8 9	Length 780;
	Best Local Similarity	22.0%;	Pred. No. 10;		
Matches	35; Conservative	29; Mismatches	80; Indels	15; Gaps	6
OY	10 CNKTEKKMMVEVDYTRKPRIILASGSV-ASIKRSNLCH---- <td>64</td> <td></td> <td></td> <td></td>	64			
	: : : : : : :	:	:	:	:
Db	282 GENPEFENNEVTSCFMYPGLPDLTLNVRIRIVKNCKLQEDDKLTMSNSQSMSGKR-	340			
OY	65 LNFVDPSPILPONVITLYTDEDEWSS--WNKIGDPIVLIELRRADVMIIAPLSANTLAIRA	122			
	: : : : : : :	:	:	:	:
Db	341 -NSTGNMNVLDIPTCVSDYSMTSTCEMKMGNGTNSTELRLLYLQLVFLLSEAHTCIPENN	399			
OY	123 GG---LCDNILTCIVRADYISKPLFLVALPAAMNTLMNNNP	158			
	: : : : : : :	:	:	:	:

Db 400 GGAGCVCHLLMDVVSADNYTLIDLW---AGQQLMKGSF 435

RESULT 9
US-09-935-868-38

; Sequence 38, Application US/09935868
; Patent No. US20020164690A1

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc

; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using

; FILE REFERENCE: REG 203D

; CURRENT APPLICATION NUMBER: US/09/935,868

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: PCT/US99/22045

; PRIOR FILING DATE: 1999-09-22

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 38

; LENGTH: 780

; TYPE: PRT

; ORGANISM: Homo Sapiens

US-09-935-868-38

Query Match 6.2%; Score 71.5; DB 9; Length 780;

Best Local Similarity 22.0%; Pred. No. 10;

Matches 35; Conservative 29; Mismatches 80; Indels 15; Gaps 6;

OY 10 CNKIEKKMNEVDYTRKRRIILAAAGSV-ASIKFSNLCH---CFSENAEYKAVASKSS 64

Db 282 CENPEERENVNTSCFMVPGVLPDNLNTVIRIKRKIKLCYEDDKLWSNMSQESIGKKR- 340

OY 65 LNFVDRPSLPQWNTLYTDEDEMS--WNKIGDPVLIHIELRRADVMIILPLSANTLAKIA 122

Db 341 -NSTGMKAVLQEPCTCYSDYMSISTCEKMKNGPTNCSTELRLYLQVFLFLESEATTCIPENN 399

OY 123 GG---LCDNLITCIVRAMDYSKPLFVAPAMNTLMNNPF 158

Db 400 GGAGCVCHLLMDVVSADNYTLIDLW---AGQQLMKGSF 435

RESULT 10

US-09-935-868-42

; Sequence 42, Application US/09935868

; Patent No. US20020164690A1

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc

; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using

; FILE REFERENCE: REG 203D

; CURRENT APPLICATION NUMBER: US/09/935,868

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: PCT/US99/22045

; PRIOR FILING DATE: 1999-09-22

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 42

; LENGTH: 780

; TYPE: PRT

; ORGANISM: Homo Sapiens

US-09-935-868-42

Query Match 6.2%; Score 71.5; DB 9; Length 780;

Best Local Similarity 22.0%; Pred. No. 10;

Matches 35; Conservative 29; Mismatches 80; Indels 15; Gaps 6;

OY 10 CNKIEKKMNEVDYTRKRRIILAAAGSV-ASIKFSNLCH---CFSENAEYKAVASKSS 64

Db 282 CENPEERENVNTSCFMVPGVLPDNLNTVIRIKRKIKLCYEDDKLWSNMSQESIGKKR- 340

OY 65 LNFVDRPSLPQWNTLYTDEDEMS--WNKIGDPVLIHIELRRADVMIILPLSANTLAKIA 122

Db 341 -NSTGMKAVLQEPCTCYSDYMSISTCEKMKNGPTNCSTELRLYLQVFLFLESEATTCIPENN 399

OY 123 GG---LCDNLITCIVRAMDYSKPLFVAPAMNTLMNNPF 158

Db 400 GGAGCVCHLLMDVVSADNYTLIDLW---AGQQLMKGSF 435

Db 400 GGAGCVCHLLMDVVSADNYTLIDLW---AGQQLMKGSF 435

RESULT 11
US-09-935-868-48

; Sequence 48, Application US/09935868
; Patent No. US20020164690A1

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc

; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using

; FILE REFERENCE: REG 203D

; CURRENT APPLICATION NUMBER: US/09/935,868

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: PCT/US99/22045

; PRIOR FILING DATE: 1999-09-22

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 48

; LENGTH: 782

; TYPE: PRT

; ORGANISM: Homo Sapiens

US-09-935-868-48

Query Match 6.2%; Score 71.5; DB 9; Length 782;

Best Local Similarity 22.0%; Pred. No. 10;

Matches 35; Conservative 29; Mismatches 80; Indels 15; Gaps 6;

OY 10 CNKIEKKMNEVDYTRKRRIILAAAGSV-ASIKFSNLCH---CFSENAEYKAVASKSS 64

Db 282 CENPEERENVNTSCFMVPGVLPDNLNTVIRIKRKIKLCYEDDKLWSNMSQESIGKKR- 340

OY 65 LNFVDRPSLPQWNTLYTDEDEMS--WNKIGDPVLIHIELRRADVMIILPLSANTLAKIA 122

Db 341 -NSTGMKAVLQEPCTCYSDYMSISTCEKMKNGPTNCSTELRLYLQVFLFLESEATTCIPENN 399

OY 123 GG---LCDNLITCIVRAMDYSKPLFVAPAMNTLMNNPF 158

Db 400 GGAGCVCHLLMDVVSADNYTLIDLW---AGQQLMKGSF 435

RESULT 12

US-09-935-868-52

; Sequence 52, Application US/09935868

; Patent No. US20020164690A1

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc

; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using

; FILE REFERENCE: REG 203D

; CURRENT APPLICATION NUMBER: US/09/935,868

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: PCT/US99/22045

; PRIOR FILING DATE: 1999-09-22

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 52

; LENGTH: 782

; TYPE: PRT

; ORGANISM: Homo Sapiens

US-09-935-868-52

Query Match 6.2%; Score 71.5; DB 9; Length 782;

Best Local Similarity 22.0%; Pred. No. 10;

Matches 35; Conservative 29; Mismatches 80; Indels 15; Gaps 6;

OY 10 CNKIEKKMNEVDYTRKRRIILAAAGSV-ASIKFSNLCH---CFSENAEYKAVASKSS 64

Db 282 CENPEERENVNTSCFMVPGVLPDNLNTVIRIKRKIKLCYEDDKLWSNMSQESIGKKR- 340

OY 65 LNFVDRPSLPQWNTLYTDEDEMS--WNKIGDPVLIHIELRRADVMIILPLSANTLAKIA 122

Db 341 -NSTGMKAVLQEPCTCYSDYMSISTCEKMKNGPTNCSTELRLYLQVFLFLESEATTCIPENN 399

OY 123 GG---LCDNLTCIVRAMDYSKPLFVAPAMNTLMNNPF 158
Db 400 GGAGCVCHLMDVVSADNTLIDLW---AGQOLLMKGSF 435

RESULT 13

US-10-109-886-8
; Sequence 8, Application US/10109886
; Patent No. US20020119499A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; TITLE OF INVENTION: ANTAGONIST TO PPAR
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/10/109,886
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-10-109-886-8

Query Match 6.2%; Score 70.5; DB 12; Length 2441;
Best Local Similarity 21.1%; Pred. No. 64;
Matches 54; Conservative 32; Mismatches 91; Indels 79; Gaps 13;

OY 10 CCKIEKKMMNEVDYTR-----KPRLLAASGSVASI-----KFSNLCH-- 48
Db 1179 CCKLAIEFEBEIDPVWQSLGCGCKRKEFSPQTLCCYKQCLTIPRDAAYSYQNKYHFC 1238
OY 49 --CFSEMAEYKAVASKSLNFVDKPSLPQNVTLTYDEDEWSSMNKIGDPVLIHELRRMAD 106
Db 1239 GKCFTF-----IQGENVTGLDDPSQPQ--TTISKQDFE--KKKNDTLDEPPVDCKECGR 1289
OY 107 VM-----IIAPLSANTLAKIAGCLDNLTCTIVRAMDYSKPLFVAPAMNTLMNNP 157
Db 1290 KMHQICVLHYDIIMP-----SGFVCDNCLKTKGRPRKKNK--FSAKRLOTTLRGNH 1338
OY 158 FTER-----HLVLLDELGITLPIPIKKLACGDYNGAMAEPSLIYS 199
Db 1339 LEDRVNKKFLRRONHPEAGEVFYRVVASSDKTVEVKPKMKSRFV---DSGEMSE--SEPYR 1393
OY 200 TVRLFWESQARKORDG 215
Db 1394 TKALF---AFEEIDG 1405

RESULT 14

US-10-109-886-10
; Sequence 10, Application US/10109886
; Patent No. US20020119499A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; TITLE OF INVENTION: ANTAGONIST TO PPAR
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/10/109,886
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734

; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 2442
; TYPE: PRT
; ORGANISM: human
US-10-109-886-10

Query Match 6.2%; Score 70.5; DB 12; Length 2442;
Best Local Similarity 21.1%; Pred. No. 64;
Matches 54; Conservative 32; Mismatches 91; Indels 79; Gaps 13;

OY 10 CCKIEKKMMNEVDYTR-----KPRLLAASGSVASI-----KFSNLCH-- 48
Db 1178 CCKLAIEFEBEIDPVWQSLGCGCKRKEFSPQTLCCYKQCLTIPRDAAYSYQNKYHFC 1237
OY 49 --CFSEMAEYKAVASKSLNFVDKPSLPQNVTLTYDEDEWSSMNKIGDPVLIHELRRMAD 106
Db 1238 EKCFTF-----IQGENVTGLDDPSQPQ--TTISKQDFE--KKKNDTLDEPPVDCKECGR 1288
OY 107 VM-----IIAPLSANTLAKIAGCLDNLTCTIVRAMDYSKPLFVAPAMNTLMNNP 157
Db 1289 KMHQICVLHYDIIMP-----SGFVCDNCLKTKGRPRKKNK--FSAKRLOTTLRGNH 1337
OY 158 FTER-----HLVLLDELGITLPIPIKKLACGDYNGAMAEPSLIYS 199
Db 1338 LEDRVNKKFLRRONHPEAGEVFYRVVASSDKTVEVKPKMKSRFV---DSGEMSE--SEPYR 1392
OY 200 TVRLFWESQARKORDG 215
Db 1393 TKALF---AFEEIDG 1404

RESULT 15
US-09-870-162A-9
; Sequence 9, Application US/09870162A
; Patent No. US20020042118A1
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica
; FILE REFERENCE: BC1006 US DIV
; CURRENT APPLICATION NUMBER: US/09/870,162A
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 09/516914
; PRIOR FILING DATE: 2000-03-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Thauera aromatica
US-09-870-162A-9

Query Match 6.1%; Score 70; DB 10; Length 194;
Best Local Similarity 34.2%; Pred. No. 2.3;
Matches 25; Conservative 10; Mismatches 30; Indels 8; Gaps 3;

OY 108 MIAPLSANTLAKIAGCLDNLTCTIVRAMDYSKPLFVAPAMNTLMNNPFTRHVLVLL- 166
Db 80 MIAPCSIKTILSAVANSFNTNLL---IRADVA--LKERKKVLMREPTLHGHRLMT 134
OY 167 --DELGITLPIPI 177
Db 135 QATENGAVLLPL 147

Search completed: November 25, 2002, 10:19:30
Job time : 12 secs

NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-09-632-711-52

Query Match 16.9%; Score 194; DB 4; Length 444;
Best Local Similarity 31.1%; Pred. No. 2,6e-15;
Matches 56; Conservative 35; Mismatches 71; Indels 18; Gaps 6;

OY 29 RLLASGSVASIKFSNLCCHCFSE-WAEYKAVASKSLNFVDPKPSLPQWVTLTYTDEMS 87
DB 31 KIVLGVSGGIAAYKTPELVRLDRGADVAVMTEAKAFITPLSL-QAVSGTPVSD--- 86
OY 88 SWNKIGDPVL-----HIEL-----RRWADVMIAPLSANTLARIAGLCNLTCTIVRAND 138
DB 87 ---SLDDPAEAAMGHIELGXXXXXKWADVLILPATADLIARVAAGMANDIVSTICLATP 143
OY 139 YSKPLFVAPAMNTLMWNNFTERRHL-VLDELGITLPIPIKRLACGDYNGAMAEPSLI 197
DB 144 XXAPAVLIPAMNQMYRAAATQHNLEVLAXSRGLLWGPDSGSGQACDIPGRXXDPLTI 203

RESULT 7

US-09-632-703B-52
Sequence 52, Application US/09632703B
Patent No. 6379553

GENERAL INFORMATION:

APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunnet, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/632,703B

FILING DATE: 24-Aug-2000

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/822,774

FILING DATE: 21-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: BARKER, M. Paul

REGISTRATION NUMBER: 32,013

REFERENCE/DOCKET NUMBER: 4121,0116-04

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 849-6613

TELEFAX: (650) 849-6666

INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:

LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-09-632-703B-52

Query Match 16.9%; Score 194; DB 4; Length 444;
Best Local Similarity 31.1%; Pred. No. 2,6e-15;
Matches 56; Conservative 35; Mismatches 71; Indels 18; Gaps 6;

OY 29 RLLASGSVASIKFSNLCCHCFSE-WAEYKAVASKSLNFVDPKPSLPQWVTLTYTDEMS 87
DB 31 KIVLGVSGGIAAYKTPELVRLDRGADVAVMTEAKAFITPLSL-QAVSGTPVSD--- 86
OY 88 SWNKIGDPVL-----HIEL-----RRWADVMIAPLSANTLARIAGLCNLTCTIVRAND 138
DB 87 ---SLDDPAEAAMGHIELGXXXXXKWADVLILPATADLIARVAAGMANDIVSTICLATP 143
OY 139 YSKPLFVAPAMNTLMWNNFTERRHL-VLDELGITLPIPIKRLACGDYNGAMAEPSLI 197
DB 144 XXAPAVLIPAMNQMYRAAATQHNLEVLAXSRGLLWGPDSGSGQACDIPGRXXDPLTI 203

RESULT 8

US-09-632-702-52
Sequence 52, Application US/09632702
Patent No. 6444428

GENERAL INFORMATION:

APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Purifying and Identifying Same

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: David J. Kulik, Evenson, McKee, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/632,702

FILING DATE: 04-Aug-2000

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/822,774

FILING DATE: 21-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: KULIK, David J.

REGISTRATION NUMBER: 36,576

REFERENCE/DOCKET NUMBER: 1486/43163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-8800

TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 52:

SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:


```

;
; ORGANISM: Escherichia coli
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-09-632-702-52

Query Match          16.9%; Score 194; DB 4; Length 444;
Best Local Similarity 31.1%; Pred. No. 2.6e-15;
Matches 56; Conservative 35; Mismatches 71; Indels 18; Gaps 6;

OY 29 RLLAASGSVASIKFSNLCHESE-WAEKAVAKSSSLNFVDKPSLPONTLYTDEDEMS 87
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 31 KIVLGSGLIAATKPELVRLRDRGADVAVMTEAKAFITPLSL-QAVSGTPVSD--- 86
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
OY 88 SMNKIGDPVL-----HIEL-----RRADYMIAPLSANTLAKIAGCLDNLTICIVRAMD 138
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 87 ---SLIDPAEAMGHIEGXXXXXKWDVLIAPATADILARVAAGMADNLSTICLATP 143
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
OY 139 YSKPLFVAAMNTLMNNPFTERRHL-VLLDELGITLIPPIKKKLAGDGNMGMAPPSUL 197
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 144 XAPVAVLPAAMNQOMRAAATOHNLVLAAXSRGLIWDGSGQACGDIGPGRXXDPLTI 203
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 9
US-09-627-376-9
; Sequence 9, Application US/09627376
; Patent No. 6342385
; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia          Caulfield, Page          Chen, Ping
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: UAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 188
; TYPE: PPT
; ORGANISM: Streptococcus mutans
US-09-627-376-9

Query Match          16.5%; Score 188.5; DB 4; Length 188;
Best Local Similarity 28.2%; Pred. No. 3.3e-15;
Matches 42; Conservative 35; Mismatches 63; Indels 9; Gaps 4;

OY 19 MEVDVTRRPRILLASGSVASIKFSNLCHESEMAEVAKAVAKSSSLNFVDKPSLPONV- 77
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 1  MEQNIIEKKILCLTGSGLLGIA-EYITFLVRFKHVRIVYSDNAKMLPVAIIQDCE 59
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
OY 78 TLYTDEDEMSMNKIGDPVLHIELRRADYMIAPLSANTLAKIAGCLDNLTICIVRAM 137
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 60 KYTTEVESTDKOK-----NHIALTRMADITVLPATANIGKANGIADNEMTTLLS- 113
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
OY 138 DYSKPLFVAAMNTLMNNPFTERRHLVL 166
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 114 -SSKPLVLTFCMNMNMENPVYQKNEVL 141
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 10
US-08-822-774-19
; Sequence 19, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
```

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
US-08-822-774-19

Query Match          14.1%; Score 162; DB 4; Length 403;
Best Local Similarity 27.1%; Pred. No. 2.1e-11;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;

OY 23 TYVRK---PRILLAASGSVASIKFSNLCHESEMAEVAKAVAKSSSLNFVDKPSLPONTVL 79
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 11  TKSRLVGRKRIYXXPSIALDVKACEGLIRHGAEVHVMSEAAKRIHP----- 61
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
OY 80 YTDEDEMSMN-KIGDPVL-----HIEL-----RRADYMIAPLSANTLAKIAGCL 126
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 62 -----YAWNLPGRNPVLTETITGTEHVELAGEHKKADLLVCAANTISKIAGCID 114
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
OY 127 DNLTICIVRAMDYSKPLFVAAMNTLMNNPFTERRHLVLLDELGITLIPP 176
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 115 DPLVTIVVTAPRPHIIMTAPAMHETMYRRPIVRENIERKKRGVFEIG 164
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 11
US-09-632-711-19
; Sequence 19, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
```

REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-632-711-19

Query Match 14.1%, Score 162; DB 4; Length 403;
Best Local Similarity 27.1%; Pred. No. 2,16-11;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;

OY 23 TVTRK---PRILLASGVASIKFSNLCHCFSEMAEYKAIVASKSSLNVDRPSLPQNTL 79
| : | | : | : | : | : | : | : | : | : | :
Db 11 TKSRRLVGKKITXPPGSIADLVKACGLIRHGAEVHAVMSEAFTKIHP----- 61

OY 80 YTDDEMSWN-KIGDPVL-----HIET----RRADVMIIAPLSANTLAKIAGCLC 126
| : | : | : | : | : | : | : | : | : | : | :
Db 62 -----YANMLPGNPVPITEGTGTGEHVLLAGEHNKANADLTVCAPAINTSKTAGCID 114

OY 127 DNLTLCTIVRANDYSKPFLVAPAMNLTMMNNPETERHLVLDLGITLIIP 176
| : | : | : | : | : | : | : | : | : | : | :
Db 115 DTPVTTVTTAFPHPIIMAIAPAHMHTMYRHPIRVRENIERLKTKGVEIFGP 164

RESULT 12
US-09-632-703B-19
Sequence 19, Application US/09632703B
Patent No. 6379553
GENERAL INFORMATION:
APPLICANT: HOGREFE, HOLLY
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finegan, Henderson, Farbow, Garrett, & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: <unknown>
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,703B
FILING DATE: 24-Aug-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/822,774
FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BARKER, M. Paul
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 4121,0116-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 849-6613
TELEFAX: (650) 849-6666
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein

```

:      HYPOTHETICAL: NO  

:      SEQUENCE DESCRIPTION: SEQ ID NO: 19:  

US-09-632-703B-19  

Query Match          14.1%; Score 162; DB 4; Length 403;  

Best Local Similarity 27.1%; Pred. No. 2.1e-11;  

Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;  

QY 23 TYTRK---PRILLASGSVASIKFSNLCHCFBSMAVKAIVASKSSLNFPVDKPSLPQVTL 79  

| : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  

Db 11 TKSRRLVGKKIYXXXXPGSIALDVKACEGLIRHGAEVHAWSEAAKIHP----- 61  

QY 80 YTDEDESSWN-KIGDPVL-----HTEL----RRADVMIIAPLSANTLAKIAGLC 126  

| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  

Db 62 -----YAMNLEPTGNPVYTELTGFEHYELAGEHNKADLLVCATANTISKIACGD 114  

QY 127 DNLTCIVRAMDYSKFLFVPARAMNTLMNNPTEHLVLDELGITLIP 176  

| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  

Db 115 DTPVTVVTTAPPRIPIIMIAPIAMHEMTYRHRPIRENIERLKRGVEFIGP 164  

RESULT 13  

US-09-632-702-19  

: Sequence 19, Application US/09632702  

: Patent No. 6444428  

GENERAL INFORMATION:  

APPLICANT: HOGREFE, HOLLY  

TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  

Extracts, PEF Protein Complexes, Isolated PEF Proteins  

and Methods for Purifying and Identifying Same  

NUMBER OF SEQUENCES: 61  

CORRESPONDENCE ADDRESS:  

ADDRESSSEE: David J. Kulik, Evenson, McKeown, Edwards &  

Leahman, P.L.L.C.  

STREET: 1200 G Street, N.W., Suite 700  

CITY: Washington  

STATE: D.C.  

ZIP: 20005  

COMPUTER READABLE FORM:  

MEDIUM TYPE: Floppy disk  

COMPUTER: IBM PC compatible  

OPERATING SYSTEM: PC-DOS/MS-DOS  

SOFTWARE: PatentIn Release #1.0, Version #1.25  

CURRENT APPLICATION DATA:  

APPLICATION NUMBER: US/09/632, 702  

FILING DATE: 04-Aug-2000  

CLASSIFICATION: <Unknown>  

PRIOR APPLICATION DATA:  

APPLICATION NUMBER: US/08/822,774  

FILING DATE: 21-MAR-1997  

ATTORNEY/AGENT INFORMATION:  

NAME: KULIK, David J.  

REGISTRATION NUMBER: 36,576  

REFERENCE/DCKET NUMBER: 1486/43163  

TELECOMMUNICATION INFORMATION:  

TELEPHONE: (202) 628-8800  

TELEFAX: (202) 628-8844  

INFORMATION FOR SEQ ID NO: 19:  

SEQUENCE CHARACTERISTICS:  

LENGTH: 403 amino acids  

TYPE: amino acid  

TOPOLOGY: unknown  

MOLECULE TYPE: protein  

HYPOTHETICAL: NO  

SEQUENCE DESCRIPTION: SEQ ID NO: 19:  

US-09-632-702-19  

Query Match          14.1%; Score 162; DB 4; Length 403;  

Best Local Similarity 27.1%; Pred. No. 2.1e-11;  

Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;  

QY 23 TYTRK---PRILLASGSVASIKFSNLCHCFBSMAVKAIVASKSSLNFPVDKPSLPQVTL 79  

| : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 10:10:25 ; Search time 36 Seconds

(without alignments)
803.205 Million cell updates/sec

Title: US-09-868-300-8

Perfect score: 1145
Sequence: 1 EFGTSSFLGCKIKKKMMME.....YSTVRLWESQARKRODCTS 217

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1145	100.0	217	21	AAV96816
2	946	82.6	209	21	AAAG11381
3	943	82.4	199	21	AAAG11382
4	449	39.2	309	22	ABBI1631
5	449	39.2	309	22	AAAM1256
6	445	38.9	204	22	AAAB94179
7	444	38.8	204	22	AAAB94470
8	443	38.7	326	21	AAV32199
9	422	36.9	170	22	AAU18952
10	422	36.9	170	22	AAU17967

11	274.5	24.0	181	23	ABP28494
12	243.5	21.3	178	23	ABBS3882
13	243.5	21.3	180	23	ABP28493
14	242.5	21.2	399	22	AAAG2741
15	242.5	21.2	400	23	ABP39940
16	227.5	19.9	399	23	ABBA4692
17	221	19.3	228	19	AAW98418
18	220	19.2	422	22	AAW99446
19	214	18.7	420	22	AAAG91520
20	204.5	17.9	181	24	AAAG3345
21	202.5	17.7	181	20	AAV43439
22	195	17.0	269	22	AAU63490
23	188.5	16.5	188	23	ABBO9642
24	169	14.8	401	19	AAAB9653
25	163.5	14.3	403	19	AAAB2844
26	141	12.3	79	23	ABP08086
27	86	7.5	238	22	ABG07040
28	85.5	7.5	669	22	ABAB7461
29	84.5	7.4	280	22	ABBB0590
30	84.5	7.4	1717	22	ABG20672
31	84	7.3	197	22	AAAG91615
32	83.5	7.3	192	20	AAV34858
33	83.5	7.3	505	22	AAAB95727
34	82.5	7.2	524	22	AAAB92814
35	82	7.2	159	22	AAAG81370
36	82	7.2	329	19	AAAB98743
37	82	7.2	418	22	AAAG82097
38	82	7.2	431	23	ABP38206
39	82	7.2	657	22	AAAB39041
40	82	7.2	996	22	ABG30156
41	81	7.1	192	20	AAV36979
42	80.5	7.0	475	22	AAAB93993
43	80.5	7.0	475	22	AAAB92735
44	80.5	7.0	521	22	ABBS8663
45	80.5	7.0	686	22	AAAB95291

ALIGNMENTS

RESULT 1
AAV96816
ID AAV96816 standard; Protein; 217 AA.
XX
AC AAV96816;
XX
DT 26-SEP-2000 (first entry)
XX
DE A. thaliana VB89 (HML3) CDC2b interacting protein.
XX
DE Cyclin-dependent protein kinase; CDK; CDC2a; CDC2b; VB89; HML3;
KW cell cycle; interacting protein; environmental stress; growth regulator;
KW herbicide; nematode resistance; plant breeding.
XX
OS Arabidopsis thaliana.
XX
PN WO200036124-A2.
XX
PD 22-JUN-2000.
XX
PF 17-DEC-1999; 99WO-EP10084.
XX
PR 17-DEC-1998; 98EP-0124062.
XX
PA (CROP-) CROPESTGN NV.
XX
PI De Veylder L, Boudolf VKCK, Torres Acosta JA, Inze D;
XX WPI: 2000-431601/37.
XX N-PSDB; AAA51412.
XX Nucleic acids encoding plant cell cycle interacting proteins, useful
PT for regulating plant growth and in recombinant DNA protocols

XX Claim 1; page 126; 152pp; English.
PS
XX The Vb89 clone encodes the Arabidopsis thaliana HAL3 homologue, a
CC halotolerant gene isolated in Saccharomyces cerevisiae. The Vb89 clone
CC interacts with A. thaliana CDC2b (a cyclin-dependent protein kinase
CC (CDK)), but not with CDC2a in the two-hybrid system.
CC CDC2a and CDC2b are the only CDK genes to have been characterized in
CC detail in Arabidopsis thaliana. They were used as bait in a two-hybrid
CC screening assay with a cDNA library of a plant cell suspension as prey.
CC The plant cell cycle interacting proteins identified were designated
CC LDV15, PH80-like protein, Vb33, Vb89, VbDHP and VbHSF. The nucleic
CC acids, vectors comprising them, the proteins they express, antibodies
CC that bind to them and or inhibitors of their protein expression and/or
CC activity may be used for modulating the cell cycle in an animal or plant,
CC plant cell division and/or growth, for influencing the activity of cell
CC cycle proteins in a plant or animal cell, as positive or negative
CC regulators of cell proliferation, for modifying the growth inhibition
CC caused by environmental stress conditions (e.g. to improve growth of
CC plants in normal or suboptimal nutrient conditions, especially
CC phosphorus), for use in a screening method for inhibitors or activators
CC of cell cycle protein, as growth regulators, herbicides and/or for
CC inducing nematode resistance in plants. The DNA sequences and their
CC regulatory sequences may be used as markers in plant or animal cell and
CC tissue cultures or as a marker in marker-assisted plant breeding. The
CC regulatory sequences may also be used for the expression of heterologous
CC DNA sequences during a stage of the cell cycle.
SQ
XX Sequence 217 AA:
Query Match 100.0%; Score 1145; DB 21; Length 217;
Best Local Similarity 100.0%; Pred. NO. 1.6e-120;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ERTGSFELCNTEKKNEVDVTFRKPRILAAAGSVASIKRNSNCHCFSEAEKAVA 60
DB 1 ERTGSFELCNTEKKNEVDVTFRKPRILAAAGSVASIKRNSNCHCFSEAEKAVA 60
OY 61 SKSLSNFVDPKSLPONVLTLYTDEDEMSWNKIGDPVLAHELRRADVMTIAPLSANTLAK 120
DB 61 SKSLSNFVDPKSLPONVLTLYTDEDEMSWNKIGDPVLAHELRRADVMTIAPLSANTLAK 120
OY 121 IAGGLCDNLITCTIVRAMDYSKPLFVAPAMNTLMNNPTEFRLHVLDELGITLPIPIKKK 180
DB 121 IAGGLCDNLITCTIVRAMDYSKPLFVAPAMNTLMNNPTEFRLHVLDELGITLPIPIKKK 180
OY 181 LACGDYNGAMAEPSLIYSTVRLFWESQARKORDGTS 217
DB 181 LACGDYNGAMAEPSLIYSTVRLFWESQARKORDGTS 217
RESULT 2
AAG11381
ID AAG11381 standard: Protein; 209 AA.
XX
AC AAG11381;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10076.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN Ep1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

[illegible]

PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158202
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	14-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159331
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159684
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160817
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	23-OCT-1999	9905-0161404
PR	23-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161922
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

Query Match	82.6%	Score 946;	DB 21;	Length 209;
Best Local Similarity	85.6%	Pred. NO. 4.1e-98;		
Matches 173; Conservative	17;	Mismatches 12;	Indels 0;	Gaps 0;

[illegible]

RESULT 3	
AAg11382	
ID	AAg11382 standard; Protein; 199 AA.
XX	
XX	
AC	AAg11382;
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 10077.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
XX	
EP1033405-A2	
NN	

XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990US-0121825.
PR 05-MAR-1999; 990US-0123180.
PR 09-MAR-1999; 990US-0123548.
PR 23-MAR-1999; 990US-0125788.
PR 25-MAR-1999; 990US-0126264.
PR 29-MAR-1999; 990US-0126785.
PR 01-APR-1999; 990US-0127462.
PR 06-APR-1999; 990US-0128234.
PR 08-APR-1999; 990US-0128714.
PR 16-APR-1999; 990US-0129845.
PR 19-APR-1999; 990US-0130077.
PR 21-APR-1999; 990US-0130449.
PR 23-APR-1999; 990US-0130510.
PR 23-APR-1999; 990US-0130891.
PR 28-APR-1999; 990US-0131449.
PR 30-APR-1999; 990US-0132048.
PR 30-APR-1999; 990US-0132407.
PR 04-MAY-1999; 990US-0132484.
PR 05-MAY-1999; 990US-0132485.
PR 06-MAY-1999; 990US-0132486.
PR 07-MAY-1999; 990US-0132487.
PR 11-MAY-1999; 990US-0134256.
PR 14-MAY-1999; 990US-0134218.
PR 14-MAY-1999; 990US-0134219.
PR 14-MAY-1999; 990US-0134221.
PR 14-MAY-1999; 990US-0134370.
PR 18-MAY-1999; 990US-0134768.
PR 19-MAY-1999; 990US-0134941.
PR 20-MAY-1999; 990US-0135124.
PR 21-MAY-1999; 990US-0135353.
PR 24-MAY-1999; 990US-0135629.
PR 25-MAY-1999; 990US-0136021.
PR 27-MAY-1999; 990US-0136392.
PR 28-MAY-1999; 990US-0136782.
PR 01-JUN-1999; 990US-0137222.
PR 03-JUN-1999; 990US-0137528.
PR 04-JUN-1999; 990US-0137502.
PR 07-JUN-1999; 990US-0137724.
PR 08-JUN-1999; 990US-0138094.
PR 10-JUN-1999; 990US-0138540.
PR 10-JUN-1999; 990US-0138847.
PR 14-JUN-1999; 990US-0139119.
PR 16-JUN-1999; 990US-0139452.
PR 16-JUN-1999; 990US-0139453.
PR 17-JUN-1999; 990US-0139492.
PR 18-JUN-1999; 990US-0139454.
PR 18-JUN-1999; 990US-0139455.
PR 18-JUN-1999; 990US-0139456.
PR 18-JUN-1999; 990US-0139457.
PR 18-JUN-1999; 990US-0139458.
PR 18-JUN-1999; 990US-0139459.
PR 18-JUN-1999; 990US-0139460.
PR 18-JUN-1999; 990US-0139461.
PR 18-JUN-1999; 990US-0139462.
PR 18-JUN-1999; 990US-0139463.
PR 18-JUN-1999; 990US-0139750.
PR 18-JUN-1999; 990US-0139763.
PR 21-JUN-1999; 990US-0139817.
PR 22-JUN-1999; 990US-0139899.
PR 23-JUN-1999; 990US-0140353.
PR 23-JUN-1999; 990US-0140354.
PR 24-JUN-1999; 990US-0140695.
PR 28-JUN-1999; 990US-0140823.
PR 29-JUN-1999; 990US-0140991.
PR 30-JUN-1999; 990US-0141287.
PR 01-JUL-1999; 990US-0141842.
PR 01-JUL-1999; 990US-0142154.

PR 02-JUL-1999; 990US-0142055.
PR 06-JUL-1999; 990US-0142390.
PR 08-JUL-1999; 990US-0142803.
PR 09-JUL-1999; 990US-0142920.
PR 12-JUL-1999; 990US-0142977.
PR 13-JUL-1999; 990US-0143542.
PR 14-JUL-1999; 990US-0143624.
PR 15-JUL-1999; 990US-0144005.
PR 16-JUL-1999; 990US-0144085.
PR 16-JUL-1999; 990US-0144086.
PR 19-JUL-1999; 990US-0144325.
PR 19-JUL-1999; 990US-0144331.
PR 19-JUL-1999; 990US-0144332.
PR 19-JUL-1999; 990US-0144333.
PR 19-JUL-1999; 990US-0144334.
PR 19-JUL-1999; 990US-0144335.
PR 20-JUL-1999; 990US-0144352.
PR 20-JUL-1999; 990US-0144632.
PR 20-JUL-1999; 990US-0144814.
PR 21-JUL-1999; 990US-0144814.
PR 21-JUL-1999; 990US-0145086.
PR 21-JUL-1999; 990US-0145088.
PR 22-JUL-1999; 990US-0145085.
PR 22-JUL-1999; 990US-0145087.
PR 22-JUL-1999; 990US-0145089.
PR 22-JUL-1999; 990US-0145192.
PR 23-JUL-1999; 990US-0145145.
PR 23-JUL-1999; 990US-0145218.
PR 23-JUL-1999; 990US-0145224.
PR 26-JUL-1999; 990US-0145276.
PR 27-JUL-1999; 990US-0145913.
PR 27-JUL-1999; 990US-0145918.
PR 27-JUL-1999; 990US-0145919.
PR 28-JUL-1999; 990US-0145951.
PR 02-AUG-1999; 990US-0146386.
PR 02-AUG-1999; 990US-0146388.
PR 02-AUG-1999; 990US-0146389.
PR 03-AUG-1999; 990US-0147038.
PR 04-AUG-1999; 990US-0147204.
PR 04-AUG-1999; 990US-0147302.
PR 05-AUG-1999; 990US-0147192.
PR 05-AUG-1999; 990US-0147260.
PR 06-AUG-1999; 990US-0147303.
PR 06-AUG-1999; 990US-0147416.
PR 09-AUG-1999; 990US-0147493.
PR 09-AUG-1999; 990US-0147935.
PR 10-AUG-1999; 990US-0148171.
PR 11-AUG-1999; 990US-0148319.
PR 12-AUG-1999; 990US-0148341.
PR 13-AUG-1999; 990US-0148565.
PR 13-AUG-1999; 990US-0148684.
PR 16-AUG-1999; 990US-0149368.
PR 17-AUG-1999; 990US-0149175.
PR 18-AUG-1999; 990US-0149426.
PR 20-AUG-1999; 990US-0149722.
PR 20-AUG-1999; 990US-0149723.
PR 20-AUG-1999; 990US-0149929.
PR 23-AUG-1999; 990US-0149902.
PR 23-AUG-1999; 990US-0149930.
PR 25-AUG-1999; 990US-0150566.
PR 26-AUG-1999; 990US-0150884.
PR 27-AUG-1999; 990US-0151065.
PR 27-AUG-1999; 990US-0151066.
PR 27-AUG-1999; 990US-0151080.
PR 30-AUG-1999; 990US-0151303.
PR 31-AUG-1999; 990US-0151438.
PR 01-SEP-1999; 990US-0151930.
PR 07-SEP-1999; 990US-0152363.
PR 10-SEP-1999; 990US-0153070.
PR 13-SEP-1999; 990US-0153758.
PR 15-SEP-1999; 990US-0154018.
PR 16-SEP-1999; 990US-0154039.
PR 20-SEP-1999; 990US-0154779.


```
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158032.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161999.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.4%  Score 943;  DB 21;  Length 199;
Best Local Similarity 86.9%  Pred. No. 8.4e-98;
Matches 173;  Conservative 15;  Mismatches 11;  Indels 0;  Gaps 0;
```

```
QY 19 MEVDVTRKPRILLLAASSVASIKFSNLCHEFSEMAEYKAVASKSLNFVDKSLPQWNT 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MEVNTTPRKPRVLLAASSVAIKFGNLCHEFTEMAEYRAVVTKSLHFLDKLSLPQEV 60

QY 79 LYTDEDESSMNKICDPVHLIELRMDVMIAPLSANTLAKIAGGLCDNLITCIVRAMD 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 LYTDEDESSMNKICDPVHLIELRMDVMIAPLSANTLAKIAGGLCDNLITCIVRAMD 120

QY 139 YSKPLFVAPAMTLMWNNPTEHRHLYLDELGITLIPPIKKKLACGDVNGNMAEPLSLI 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 YTKPLFVAPAMTLMWNNPTEHRHLSLDELGITLIPPIKKRLACGDVNGMAEPLSLI 180

QY 199 STVRLFWESQARKORDGTS 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 STVRLFWESQAHQOTGTS 199

RESULT 4
ABBI1631
ID ABBI1631 standard; peptide: 309 AA.
XX
AC ABBI1631;
DT 11-JAN-2002 (first entry)
XX
DE Human receptor molecule homologue, SEQ ID NO:2001.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
    haematopoiesis regulation; tissue growth; immunomodulator; activin;
```

```
KW inhibit; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asplenia; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytosolic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; ant ulcer.

OS Homo sapiens.
PN WO200157188-A2.
PN 09-AUG-2001.
PD 05-FEB-2001; 2001WO-US03800.
PF 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-457740/49.
XX N-PSDB: ABA08875.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
XX treating or ameliorating a medical condition in a mammalian subject
XX e.g. arthritis and cancer -
XX
PS Claim 20; Page 222; 1963pp; English.

CC Sequences ABBI0981-ABBI2330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
```

```
XX SQ Sequence 309 AA:
Query Match 39.2%; Score 449; DB 22; Length 309;
Best Local Similarity 46.9%; Pred. No. 7.1e-42;
Matches 92; Conservative 28; Mismatches 70; Indels 6; Gaps 3;

OY 26 RKRPRLLAASGSVASIKFENLCHCFSE--WAQYKAVASKSLSNFVDRKPSLPQNVLTYTE 83
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 51 RRFHVLVGVTGSVAALKLPLVSKLIDIPGLEVAVVTTERAKHFYSPQDIP--VTLYSDA 108
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :

OY 84 DEWSSWNKIGDPVLIHLELRMDVMIAPLSANTLAKIAGLCDNLITCIVRAMDYSKPL 143
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :
DB 109 DEWEMMKSRSDPYLIHDLRRMADLLVAPLDANTLGKVASGICDNLITCVMRAMDRSKPL 168
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :

OY 144 FVAPAMNTLMNNPFTERRLVLDLDELGITLPIPIKKKLAGCDYGNGAMAEPSLIYSTVR- 202
   ||||| || : : : : : : : : : : : : : : : : : : : : : : :
DB 169 LFCPAMNTAMMEHPITAAQVDQLKAFGYVEIPCVAKKLVCGDEGLGAMAEVGTIVDKKE 228
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :

OY 203 -LFWESQARKORDGTS 217
   || : : : : : : : : : : : : : : : : : : : : : : :
DB 229 VLFQHSFGFOQSKPGIS 244

RESULT 5
AAM41256
ID AAM41256 standard; Protein; 309 AA.
AC AAM41256;
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 6187.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QH, Zhou P, Goodrich R, Drmanac RT;
DR WPI: 2001-44253/47.
DR N-PSDB; AA160412.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2: SEQ ID NO 6187; 10078pp: English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
```

```
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, cancer diagnosis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 309 AA:
Query Match 39.2%; Score 449; DB 22; Length 309;
Best Local Similarity 46.9%; Pred. No. 7.1e-42;
Matches 92; Conservative 28; Mismatches 70; Indels 6; Gaps 3;

OY 26 RKRPRLLAASGSVASIKFENLCHCFSE--WAQYKAVASKSLSNFVDRKPSLPQNVLTYTE 83
   || : : ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
DB 51 RRFHVLVGVTGSVAALKLPLVSKLIDIPGLEVAVVTTERAKHFYSPQDIP--VTLYSDA 108
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :

OY 84 DEWSSWNKIGDPVLIHLELRMDVMIAPLSANTLAKIAGLCDNLITCIVRAMDYSKPL 143
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :
DB 109 DEWEMMKSRSDPYLIHDLRRMADLLVAPLDANTLGKVASGICDNLITCVMRAMDRSKPL 168
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :

OY 144 FVAPAMNTLMNNPFTERRLVLDLDELGITLPIPIKKKLAGCDYGNGAMAEPSLIYSTVR- 202
   ||||| || : : : : : : : : : : : : : : : : : : : : : : :
DB 169 LFCPAMNTAMMEHPITAAQVDQLKAFGYVEIPCVAKKLVCGDEGLGAMAEVGTIVDKKE 228
   || : : ||||| : : : : : : : : : : : : : : : : : : : : : : :

OY 203 -LFWESQARKORDGTS 217
   || : : : : : : : : : : : : : : : : : : : : : : :
DB 229 VLFQHSFGFOQSKPGIS 244

RESULT 6
AAB94179
ID AAB94179 standard; Protein; 204 AA.
AC AAB94179;
XX
XX 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:14492.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
```


DR 15-FEB-2000 (first entry)
XX Human receptor molecule (REC) encoded by Incyte clone 2022379.
DE
XX Receptor; REC; human; diagnosis; therapy; neoplastic disorder;
KM Immunological disorder; reproductive disorder; nervous disorder;
KM gastrointestinal disorder; smooth muscle disorder;
KM musculoskeletal disorder.
XX Homo sapiens.
XX WO957270-A2.
XX 11-NOV-1999.
XX 28-APR-1999; 99WO-US09191.
XX 01-MAY-1998; 98US-0071822.
XX (INCY-) INCYTE PHARM INC.
XX Hillman JL, Bandman O, Tang YT, Yue H, Lal P, Corley NC;
PI Guejler KJ, Patterson C;
XX MPI; 2000-052971/04.
DR N-PSDB; AAZ34608.
XX
PT Novel human receptor molecules used in the diagnosis, treatment and
PT prevention of neoplastic, immunological, reproductive gastrointestinal,
PT nervous, smooth muscle and musculoskeletal disorders.
XX
XX Claim 1; Page 71-72; 94pp; English.
XX
XX The present sequence represents a human receptor molecule (REC)
XX encoded by Incyte cDNA clone 2022379. The invention provides human
XX RECs and polynucleotides which identify and encode REC, as well as
XX vectors, host cells, antibodies, agonists and antagonists. Human
XX RECs appear to play a role in neoplastic, immunological,
XX reproductive gastrointestinal, nervous, smooth muscle and
XX musculoskeletal disorders. The protein, antagonists and agonists,
XX and compositions can be used to treat: a reproductive disorder,
XX including but not limited to, prolactin production disorders,
XX infertility including tubal disease, ovulatory defects, endometriosis,
XX disruptions of the oestrous and menstrual cycles, polycystic ovary
XX syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian
XX tumours, uterine fibroids, autoimmune disorders, ectopic pregnancy,
XX teratogenesis, breast cancer, fibrocystic breast disease,
XX galactorrhoea, disruptions of spermatogenesis, abnormal sperm
XX physiology, testis cancer, prostate cancer, benign prostatic
XX hyperplasia, prostatitis, Peyronie's disease, male breast carcinoma
XX and gynecomasia; gastrointestinal disorders including, but are not
XX limited to, dysphagia, peptic oesophagitis, oesophageal spasm and
XX stricture, oesophageal carcinoma, dyspepsia, indigestion, gastritis,
XX gastric carcinoma, anorexia, nausea, emesis, gastroparesis, intestinal
XX tract infection, peptic ulcer, colitis, Whipple's disease, Mallory-Weiss
XX syndrome, irritable bowel syndrome, short bowel syndrome, diarrhoea,
XX constipation, cirrhosis, jaundice, and hepatic vein thrombosis; nervous
XX disorders including, but are not limited to, Alzheimer's disease,
XX amnesia, bipolar disorder, cataplexy, cerebral neoplasms, Down's
XX syndrome, and dystonia; smooth muscle cell disorders including, but
XX not limited to, angina, anaphylactic shock, arrhythmia, cardiovascular
XX shock, migraine, and pheochromocytoma; musculoskeletal disorders
XX including muscular dystrophy, central core disease, nemaline myopathy,
XX centronuclear myopathy, lipid myopathy, inclusion body myositis,
XX thyrotoxic myopathy, and ethanol myopathy; immunological disorders
XX including AIDS, Addison's disease, adult respiratory distress
XX syndrome, allergy, ankylosing spondylitis, amyloidosis, anaemia,
XX asthma, atherosclerosis, autoimmune hemolytic anaemia, autoimmune
XX thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
XX disease, atopic dermatitis, dermatomyositis, diabetes mellitus,
XX emphysema, episodic lymphopenia with lymphocytotoxic, erythroblastosis
XX fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
XX Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,

CC hyperoesinophilia, irritable bowel syndrome, multiple sclerosis,
CC myasthenia gravis, myocardial or pericardial inflammation,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's
CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic
CC sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis,
CC Werner syndrome, complications of cancer, haemodialysis, and
CC extracorporeal circulation; viral, fungal, bacterial or protozoan
CC infections; trauma; and neoplastic disorders including adenocarcinoma,
CC leukemia, melanoma, myeloma, sarcoma, and various cancers. The REC
CC polynucleotide is a source of probes and primers which bind may be used
CC to detect REC in a sample from a patient (claimed). They may also be
CC administered as part of a gene therapy regime.
XX
SQ Sequence 326 AA:
Query Match 38.7%; Score 443; DB 21; Length 326;
Best local Similarity 48.6%; Pred. No. 3,6e-41;
Matches 87; Conservative 26; Mismatches 62; Indels 4; Gaps 2;
QY 26 RKPRLLAAGSGVASIKFNSLCHCFSE--MAEYKAVASKSLNFVDKPSLPQNVLTYPDE 83
DB 16 RRFHVLVGVTSVAALKLPLVLSKLDIDGLGEVAVVTTERRAHFTSPQDIP--VILYSDA 73
QY 84 DEWSWNKIGDPYLHIELRRMADVMIIAPLSANTLAKIAGLCNDLITCIVRAMDYSKPL 143
DB 74 DEWEWKSRSDPELVHIDLRMADLLVAPLDNLTGKVASGICDNLITCIVRAMDYSKPL 133
QY 144 EVAPAMNTLMNNPFTERRHLVLDELGITLPIPKKLAGCGYGAMAPSLIYSTR 202
DB 134 LFCPAMNTAMWEHPITPAQVDOLKAFGYEIPCVAKKLVCGDEGLGAMAEGVTIDVKV 192
RESULT 9
AA018952
ID AA018952 standard; Protein; 170 AA.
XX
AC AA018952;
XX
XX 21-NOV-2001 (first entry)
XX
XX Novel lung cancer antigen, Seq ID No 64.
DE Human; lung cancer; immunosuppressive; antiarthritic; antirheumatic;
XX antiproliferative; cytostatic; cardiact; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide;
XX ophthalmological; autoimmune disease; rheumatoid arthritis; neoplasm;
XX hyperproliferative disorder; cardiovascular disorder; angiogenesis;
XX nervous system disorder; Alzheimer's disease; infection; skin aging;
XX ocular disorder; wound healing; organ transplantation.
XX
OS Homo sapiens.
XX
XX WO200155300-A2.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01238.
PF
XX 31-JAN-2000; 2000US-0179065.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
PI
XX MPI; 2001-465565/50.
DR N-PSDB; AAS30612.
XX
PT Isolated nucleic acid molecule encoding a lung cancer antigen is used
PT in preventing, treating or ameliorating a medical condition
XX
XX Claim 1; SEQ ID No 64; 475pp; English.

PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476224/51.
DR N-PSDB; AAS28151.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
PS Claim 11; SED ID No 585; 546pp; English.
XX
CC The present invention relates to the isolation of novel human
CC respiratory antigens, and cDNA (AAS27869-AAS28159) and genomic
CC sequences encoding for these polypeptides. The sequences of the

CC Invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences
CC of the invention are useful in gene therapy and antisense therapy.
CC AAU17685-AAU1975 represent novel human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 170 AA:
XX
Query Match 36.9%; Score 422; DB 22; Length 170;
Best Local Similarity 52.3%; Pred. No. 3 5e-39;
Matches 81; Conservative 20; Mismatches 50; Indels 4; Gaps 2;
OY 55 EYKAVASKSLNPFVKPSLPQNTLYTDEDESSWNKIDPVLHIELRMADVMTIAPLS 114
DB 13 EVAVVTERAKHFPSPQDIP--VTLYSDADEMEMKRSRSDPVLHIDLRWADLLVAPLD 70
OY 115 ANTLAKIAGGLCDNLTECTVRANDYSKPLFVAPAMNTLMNNPFTRRHLVLDLGITLI 174
DB 71 ANTLGRVASGICDNLTCVRAWDRSKPLFCFAPAMTAMWEHPITAOQVDLKAFGYVEI 130
OY 175 PPIKKIACGDYNGNAAPEPSLIYTVR--LFWES 207
DB 131 PCYAKKLVCGDEGLGMAEYGVTVDKYKVELRQHS 165
RESULT 11
ABP28494
ID ABP28494 standard; Protein; 181 AA.
XX
AC ABP28494;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 6164.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHTR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN69125.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 1; Page 3779; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC		streptococcus/G8S (Streptococcus agalactiae) O group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5463 sequences (SI), given in the specification.	
CC	The proteins have antibacterial and anti-inflammatory activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.	
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.	
CC		
CC		
CC		
S0	Sequence 181 AA;	
	Query Match 24.0%; Score 274.5; DB 23; Length 181;	
	Best Local Similarity 35.5%; Pred. No. 1.6e-22;	
	Matches 65; Conservative 37; Mismatches 68; Indels 13; Gaps 4	
OY	30 ILLAAGSVASIKFKNLCHCFSEMA-EKKAVASKSLSNVDKPS--LFQNTLYTDEDE 85 : :: : : : : : : : : : : : : : : : : :	
Dd	5 ITLAVSGTSIAVKADLTSLQLTKIGVDHIIIMTOAQATOPTITPLTLIOVLSKNAIHLDVMDE 64 :: :	
OY	86 WSSWNRIGDP--VLHIETLRMADVMIHIAPIASNLAKIANGGLCDNLTCIVRAMDYSKPL 143 :	
Dd	65 -----HDPKYINIEELAKRTDLETVAPASANTIAHLAVGRADNLTVSVALALPATPTKK 117 :	
OY	144 FVAPEAMNTLMNNPETERHVLDELGITLPPIRKKILACGDBGYGCMAPESLIYSTVR.L 203 : :	
Dd	118 LIAPAMNTKMQNPTGTGEIRKLSTTGFEIPKSSLACGGKGREALDDIVLATIDTF 177 : :	
OY	204 FWE 206 : :	
Dd	178 IWK 180	
	RESULT 12	
ID	ABB53882 standard; Protein; 178 AA.	
XX		
AC	ABB53882;	
XX		
DT	16-MAY-2002 (first entry)	
XX		
DE	Lactococcus lactis protein dfpa.	
XX		
KW	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.	
OS	Lactococcus lactis IL1403.	
PX		
PN	FR2807446-A1.	
XX		
PD	12-OCT-2001.	
XX		
FZ	11-Apr-2000; 2000FR-0004630.	
XX		
PR	11-Apr-2000; 2000FR-0004630.	
XX		
PA	(INRG) INRA INST NAT RECH AGRONOMIQUE.	
PI	Bolotine A, Sorokline A, Renault P, Ehrlich SD;	
DR	WI; 2002-043418/06.	
XX		
PT	New nucleotide sequence useful in the identification of Lactococcus	
XX	lactis and related species -	
XX		
CS	Claim 6; SEQ ID NO 584; 2504pp; French.	
XX		

[illegible]


```
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX
XX Doucette-Stamm LA, Bush D;
PI
XX
XX WPI: 2002-381255/41.
DR N-PSDB; ABN92485.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
PS Disclosure; SEQ ID 4785; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 400 AA;

```

Query Match 21.2%; Score 242.5; DB 23; Length 400;
Best Local Similarity 32.8%; Pred. No. 1.9e-18;
Matches 61; Conservative 34; Mismatches 74; Indels 17; Gaps 4;

```
OY 30 ILLAAGSVASIKFSNL-CHCFSEMAEYKAVASKSSLNFVDKPSL-----PQNTLYTD 82
Db 5 ILLAVTGGLAAVKAIDLSKLIOSGYDVRMLSDHAEFTVPLAFOATSRNPVYNTFEK 64
OY 83 EDEMSSWNKIDPVIHILRRADVMITAPLSANTLAKIAGLCNLTCTYRANDYSKP 142
Db 65 ENP-----EEIQHVSIGDMADAIIVAPANTIAKLSVGIADDLITSTLA--TTTP 114
OY 143 LFVAPAMNTLMNNPFETRHVLVLDDELGITLPIPKKKLACG DYNGAMAEPSLITSTVR 202
Db 115 KFVAPAMVNNYNNPRTKHNNKVLSDGDIYFIEPSSGLACGYVAKGRMEPMQILSVIN 174
OY 203 LFMESQ 208
Db 175 KFTQQ 180

```

Search completed: November 25, 2002, 10:14:35
Job time : 39 secs

THIS PAGE BLANK (USPTO)